



The Normativity of Structural Rationality

Citation

Langlois, David Joseph. 2014. The Normativity of Structural Rationality. Doctoral dissertation, Harvard University.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:13067678>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

THE NORMATIVITY OF STRUCTURAL RATIONALITY

A DISSERTATION PRESENTED

BY

DAVID LANGLOIS

TO

THE DEPARTMENT OF PHILOSOPHY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN THE SUBJECT OF

PHILOSOPHY

HARVARD UNIVERSITY

CAMBRIDGE, MASSACHUSETTS

SEPTEMBER 2014

© 2014 David Langlois
All rights reserved.

THE NORMATIVITY OF STRUCTURAL RATIONALITY

ABSTRACT

Many of us take for granted that rationality requires that we have our attitudes combined only in certain ways. For example, we are required not to hold inconsistent beliefs or intentions and we are required to intend any means we see as crucial to our ends. But attempts to justify claims like these face two problems. First, it is unclear what unifies the rational domain and determines what is (and is not) rationally required of us. This is the *content* problem. Second, as philosophers have been unable to find any general reason for us to have our attitudes combined only in certain ways, it is unclear why, or in what sense, we are required to comply with these putative requirements in the first place. This is the *normativity* problem.

My dissertation offers an account of rationality which solves these problems. I argue that the entire domain of rational requirements can be derived from a single ultimate requirement demanding that we not have sets of intentions and beliefs which cause their own failure. This General Requirement of Structural Rationality explains the unity of the rational domain and directly solves the content problem. But it also solves the normativity problem. I argue that whenever we violate the General Requirement we are engaged in a form of criticizable self-undermining. I propose that this is enough to ground the claim that we ought to comply with the General Requirement's demands. This conclusion can be secured as long as we accept the thesis of normative pluralism, according to which there is more than one fundamentally distinct form of normative 'ought.'

CONTENTS

Acknowledgments	v
Chapter 1 – Two Problems About Structural Rationality	1
Chapter 2 – Constitutivism, Reasoning, and Rationality	25
Chapter 3 – Reasons Fundamental Error Theories	59
Chapter 4 – The Possibility of Normative Pluralism	85
Chapter 5 – The Content of Structural Rationality	114
Chapter 6 – The Normativity of Structural Rationality	143
Works Cited	159

ACKNOWLEDGMENTS

I'm extremely appreciative of the assistance and support I've received from the members of my dissertation committee—Tim Scanlon, Chris Korsgaard, and Selim Berker. Tim's guidance was invaluable as I prepared my prospectus and the early versions of these chapters, and he has remained a consistent source kind support throughout a challenging process. Chris' detailed (and remarkably prompt!) feedback has helped me avoid many serious mistakes and has shaped my view in many unexpected ways. Selim's influence on my dissertation and my time at Harvard has been particularly significant. For seven years, Selim has given me an unreasonable amount of his time and energy. His guidance has improved every part of this project and has made my time at Harvard far more rewarding, and fun, and fulfilling than it otherwise could have been.

I'm indebted, too, to many philosophers outside of my committee. Derek Parfit provided me with extremely detailed and thoughtful feedback at three points over the last few years, each time helping me make considerable improvements to my work. During my time in Cambridge, Peter Koellner was a good friend, a great philosophical sounding board, and an excellent flyer. Rahul Kumar is responsible (or to blame) for getting me interested in philosophy, and his friendship and mentorship has made graduate school and dissertation-writing more tolerable. Gary Francione's encouragement and friendship, and his research, have been a tremendous and positive influence on my life and work. I am thankful to Michael Kessler and Paul Schofield for organizing that Pufendorf reading group in 2010. Austin Booth has been an important ally in the battle against the darkness. Paul Julian has sent me the best links and GIFs a friend could ask for.

And I am, of course, deeply thankful to my family for their enduring support. Many

people claim to have the best parents in the world, but I'm lucky that I can say it truthfully. My mom and dad have been incredible friends and role models and, despite some poor snack-related judgments in the early days, have taken unusually good care of me over the past thirty years. I'm lucky and proud to have them as parents. My brothers, The Bluths, are my best friends, and they are some of the nicest young guys on the planet. Most of all, I am thankful to Melanie. It is due to her continuing love, and patience, and friendship that I never worry too much about how things will go from here on out.

1 – TWO PROBLEMS ABOUT STRUCTURAL RATIONALITY

1.1 – Structural Rationality and Irrationality

Imperfect agents like us often wind up with jumbles of attitudes that seem just plain wrong together. It is easy to construct true-to-life examples that illustrate the phenomenon:

(1) You plan to spend the weekend with your family while also intending to complete your new manuscript by Sunday, and while believing that doing both of these things on the same weekend is simply impossible.

(2) A friend of yours believes that her ex-husband still loves her, even while she believes this is incompatible with the mean-spirited messages he has sent her by email.

(3) I believe that I have overwhelmingly strong reason to immediately book a flight to meet my long-lost relatives, but I nevertheless have no intention to do so.

Something is going wrong with each of these agents—they are each making some kind of mistake. But, in each case, the mistake does not seem to consist in failing to do or think some discrete thing or in failing to have some particular attitude.

For example, from the brief description of your circumstances given in (1), we could not conclude that you are making a mistake by planning to finish your manuscript this weekend or, alternatively, by planning to spend the weekend with your family. Perhaps you should do only one, or perhaps you should do neither, or perhaps you could and should find

a way to do both (despite your belief). Likewise, your friend in (2) could be either right or wrong in believing that her former spouse still loves her, and could be right or wrong about the meaning of those cruel messages in her email inbox. And perhaps in (3) I would encounter life-ruining information about my family's history if I were to book airplane tickets to visit my relatives. It may be that I am achieving the best outcome by staying put at home. Nevertheless, I seem to be making some kind of mistake by not trying to book a flight while maintaining my belief.

These considerations highlight that in each of (1), (2) and (3), the agent's mistake is not located at some particular point within her or his set of attitudes. Instead, the mistake resides in the combination of attitudes as such (or, as I will sometimes say, in the agent's *arrangement* of attitudes).¹ Which particular attitudes each agent would do best to hold or abandon (if any) appears to be a different matter entirely. Indeed, the contents of and justifications for their attitudes seem to be irrelevant to the kind of error I am focused on. Any agent with an identical combination of attitude-types would be making the same mistake, regardless of the circumstances and regardless of the contents of those attitudes.²

We have a word for people who are making the kind of mistake I am describing: we call them "irrational." People with inconsistent intentions or beliefs, and people who do not endeavor to take the means to their chosen ends, and people who do not follow through on their judgments about what they should do, are all irrational. Conversely, people who do not

¹ When I say "combination of attitudes," it would be more accurate to say "combination of attitude-types." But for the sake of readability, I will most often say only "attitudes" rather than "attitude-types."

² See Brunero 2010 for a helpful discussion on this point.

have their attitudes combined in these evidently prohibited ways are (to that extent) rational.

The words “rational” and “irrational” are sometimes used to pick out other phenomena. A person might be called “irrational” for making some substantive mistake—for example, for refusing to admit that *The Wire* is a better program than *Family Ties*, or for denying humanity’s responsibility for climate change, or for failing to give sufficient weight to one’s own future interests. Such people may be irrational in some sense of the word, but they need not be irrational in the sense displayed in (1), (2), and (3). In this dissertation, I am focused narrowly on irrationality due to the structure, rather than the substance, of our attitudes. To use a bit of philosopher’s terminology, and to avoid any unnecessary confusion, we can call this phenomenon “structural irrationality” and its contrary “structural rationality.”³ Unless I say otherwise, whenever I use words like “rational” or “rationality,” I have in mind structural rationality.

1.2 – Requirements of Structural Rationality

It is common in philosophy to take for granted that there are requirements of rationality

³ The phrase “structural rationality” comes from Scanlon 2007. But the phenomenon of structural rationality has been discussed in under one label or another for decades. Bratman 1981 and Korsgaard’s “The Normativity of Instrumental Reason” (available in Korsgaard 2009a, but first published in 1997) have been consistent reference points throughout the recent debates. The most prominent and promising recent theories of this aspect of rationality have been developed by John Broome (especially in Broome 2013, but beginning in Broome 1999, Broome 2001, and Broome 2002), Niko Kolodny (especially in Kolodny 2005, Kolodny 2007, Kolodny 2008a, and Kolodny 2008b), Christine Korsgaard (especially in Korsgaard 2009a, Korsgaard 2009b, and Korsgaard 2009c), T.M. Scanlon (especially in Scanlon 1998 and Scanlon 2007), Derek Parfit (especially in Parfit 2001 and Parfit 2011a), and Joseph Raz (especially in Raz 2005a and Raz 2005b). The bulk of this dissertation is a reflection on and contribution to the ongoing discussion between these writers.

demanding that we not harbor improper combinations of attitudes. This looks like a secure and modest claim: it follows naturally from the observation that agents like those in (1), (2), and (3) are making a rational mistake simply in virtue of how they have their attitudes arranged. If agents make a rational mistake whenever they have their attitudes combined in such-and-such a way, it must be that there is a requirement of rationality demanding that we not have our attitudes so arranged. So the thought goes.

For example, corresponding to your apparent rational failure in (1), we might appeal to a requirement of rationality demanding consistency in intention. We might state the requirement as follows:

IC

You are rationally required not to [intend to X, intend to Y, and believe that you cannot both X and Y].⁴

This principle looks quite plausible. In (1) you have two intentions which you believe cannot be co-realized. You seem to be violating a demand of rationality, and this demand is expressed in IC. I will sometimes refer to these requirements of structural rationality simply as *structural requirements*.

Similarly, in characterizing your friend's rational mistake in (2), we might appeal to a requirement demanding consistency in belief. This principle could be stated like so:

BC

You are rationally required not to [believe P, believe $P \rightarrow Q$, and believe $\neg Q$].

⁴ I use square brackets to emphasize that the “rationally required” takes so-called “wide scope” over the whole set of attitudes, barring their simultaneous holding. The requirements of rationality I am putting forward here are, then, wide-scope requirements on our attitudes or states of mind. In Chapter 2, I will investigate whether this is the correct way to model the requirements of rationality. See Broome 1999, 2001, and 2002 for foundational work on wide-scope requirements of rationality.

In (2), your friend believes that her ex-spouse has sent her cruel messages and she believes that if he has sent her cruel messages then he does not love her. But she nevertheless believes that he loves her. BC states a requirement of rationality prohibiting this kind of blatant doxastic inconsistency.

Lastly, in (3), I seem to be irrational because I do not have the right kind of relationship between my belief about my reasons (for booking a flight) and my other attitudes. I am irrational because I am lacking an intention which I believe that I should hold. Recalling a Greek word pertaining to self-control, modern writers have said that rationality demands *enkrasia*⁵ in intention. Consider:

EI

You are rationally required not to [[believe that you have decisive reason to intend to X] and [not intend to X]].

This, too, has the appearance of an intuitively plausible principle. And it seems to express where I am going wrong in (3), above.

I am highlighting these three rational principles due to their relationship with my opening examples. But the domain of rational requirements is not exhausted by IC, BC, and EI. Upon reflection, it seems that there are considerably more requirements of rationality than just these. For example, one of the most widely acknowledged and discussed requirements of rationality prohibits inconsistency or incoherence between end-focused

⁵ I have been informed that the correct word is “enkrateia” rather than “enkrasia.” But to maintain consistency with most other recent work on this topic, I will use the word “enkrasia.” I will use this term when referring both to prohibitions against *akrasia* (action, intention, or belief against one’s normative judgment) and prohibitions against *accidie* (the lack of an action, intention, or belief one sees as required). I will also follow the trend of using “akrasia” to refer to both *akrasia* and *accidie*. For my purposes, these distinctions are not important.

intentions and means-focused intentions. We might state the requirement of means-end coherence like so:

ME

You are rationally required not to [intend to X, believe that you can X only if you intend to Y, and not intend to Y].⁶

It is worth pointing out that ME appears to be related to the rational principle that Kant famously called the “Hypothetical Imperative,” and to a principle which recent writers have sometimes called “The Instrumental Principle.”

The list of requirements foisted upon us by rationality does not end here, either. But I will abstain from cataloging all of the requirements of rationality at this stage.

Some of the terminology I am using is unnatural. A normal, well-adjusted person would not say to her friend, “You are rationally required not to believe P, believe $P \rightarrow Q$, and believe $\neg Q$.” A normal person would probably not even say, “You are rationally required to have consistent beliefs.” But, as the starting-point examples help to illustrate, the underlying ideas are not just philosophers’ ideas. A normal person might say, “You shouldn’t have inconsistent beliefs,” or “You’d be irrational to not take the means to your ends,” or “You must do what you think is best.” These are colloquial ways of expressing the demands of rationality.

⁶ My statement of ME is admittedly imprecise. As has been pointed out in Raz 2005a, Wallace 2001, and elsewhere, an agent who believes that she can X only if she intends to Y, but who believes that she could safely Y at some point in the future, is not necessarily irrational for failing to intend to Y right now. This is an interesting point, but one which we could easily accommodate by restating ME more precisely. (For example, we could say that you are rationally required not to [intend to X, believe that you can X only if you *now* intend to Y, and not intend to Y].) It may be that the other requirements I am considering could also be given more precise (and lengthy) statements. But it is not my goal in this project to work out fine-grained versions of every requirement of structural rationality.

1.3 – Justifying the Requirements of Rationality

That there are requirements of rationality seems undeniable. And we would not want to deny their existence anyway, as they do an excellent job of making sense of our intuitions about rationality and irrationality. But the requirements of structural rationality do not come for free, philosophically speaking. Rational requirements cannot be brute, arbitrary ordinances on the human mind. If there are indeed such things as requirements of rationality, then we must be able to explain why it is that we are required to comply with their demands.

If we have the aim of explaining the foundations of rationality, there is an obvious place to look. The requirements of rationality constrain our combinations of epistemic and practical attitudes. Perhaps, then, we can explain these principles just by citing the epistemic and practical considerations that count in favor of not holding the relevant combinations of attitudes—in virtue of which complying with the requirements of rationality is good or worthwhile. This is the flat-footed strategy of offering reasons to comply with the principles of rationality. For instance, we might propose that by complying with these principles we will gain more true beliefs, or gain fewer false beliefs, or that more of our ends will be realized, or that more total good will be produced in the world, or something else along these lines. The identified considerations, whatever they are, could make it true that we are required not to have our attitudes combined in the relevant ways.

On the kind of proposal I am considering, a claim like

You are rationally required not to [intend to X, intend to Y, and believe that you cannot both X and Y]

is alleged to be true because it is true that

You have reason not to [intend to X, intend to Y, and believe that you cannot both X and Y].

And, on this proposal, you have reason not to have your attitudes so arranged because there

is something epistemically or practically valuable about not harboring the combination of attitudes in question. This is a naïve, reasons-based account of the foundations of rational requirements.⁷

I call this approach to thinking about rationality “naïve” because, despite its simplicity and its first-blush attractiveness, it is saddled with an insuperable problem. The problem, in short, is that we do not have any sufficiently persistent and stringent reasons to comply with the principles of rationality. It is possible that compliance with the requirements of rationality is sometimes (or even often) epistemically or practically valuable or worthwhile. But it is completely implausible that compliance is always so beneficial as to ensure that we have persistent, decisive reason for compliance. For any given proposal about what generates our reasons to comply with the requirements of rationality, we can imagine counterexamples in which flouting the requirements of rationality would be better on precisely the same score. A violation of a rational requirement could, in a particular case, lead to an increase of one’s true beliefs, or an increase in one’s realized ends, or an increase in the amount of good in the world, and so on. Whatever values or goods are generally promoted by complying with the

⁷ Throughout this dissertation, I will aim to avoid getting bogged down in questions concerning the nature or structure of normative reasons. I will confine most of my comments about the nature of reasons to footnotes, in an effort to prevent the body of the project from getting pulled into a series of side debates concerning reasons.

There are a tremendous number of similar but importantly different analyses currently on offer from writers within the reasons-fundamentalist tradition, and I am hoping to avoid picking or discriminating amongst them. The most general and non-controversial description of a normative reason I can think of is this: a consideration, C, is a reason for an agent, A, to X (i.e., hold an attitude or perform an action) if, all other things being equal, C makes A’s X-ing worthwhile in some respect and to some degree.

principles listed above might in certain cases be better promoted by violating these same principles.⁸

Suppose, for example, that it is claimed that our reasons for complying with a requirement like IC are grounded in the fact that compliance assists in the realization of our ends. Now, along comes an agent whose violation of IC would, in some particular case, result in her bringing about more of her ends than she would otherwise bring about. We can assume that this would be a good outcome.⁹ However, we should be able to explain that this good outcome would be achieved *despite* the agent's failure to comply with rationality's still-persisting demand. Unfortunately, the proposal at hand is unable to accommodate this thought, because it holds that IC's force is derived from a value which, in this case, is better promoted by violating the principle. In any case where an agent's compliance with a principle like IC would not best promote the relevant value, the proponent of the proposal

⁸ I am here assuming that a reason to comply with the requirements of rationality—to not hold the attitudes that rationality prohibits—must come in the form of a benefit or value or good which would be promoted or served by complying. In other words, I am assuming that reasons to comply with rational requirements must be grounded in instrumental considerations. Arguably, not all reasons have this form. (For example, my reasons to appreciate the wilderness may not derive from some external good that is promoted by my appreciating the wilderness.) So one might wonder whether we might have what Kolodny (2005) refers to as “intrinsic” reasons to comply with rational requirements. Such a reason for rational compliance would be a consideration in virtue of which being rational is good-in-itself, even if it serves no external purpose. I am mindfully setting aside this issue until Chapter 6. As it turns out, I believe that the account of rationality which I develop in Chapter 5 provides the most plausible explanation of why we might have intrinsic reasons to be rational.

⁹ For the sake of the example, I am ignoring the fact that maximizing the realization of our ends is itself unlikely to be something which we have decisive, persistent reason to do. (For example, we may not have any reason at all to bring about our evil or malicious ends.)

will be forced to conclude that the agent is not, in that case, required to comply with the (would-be) requirement at all. This is an intuitively unacceptable result, as the agent still seems to be making a mistake by harboring the questionable combination of attitudes. A proper account of rationality must be able to explain that the end-maximizing agent with inconsistent intentions achieves a good outcome despite her violation of a rational demand.¹⁰

This observation illustrates that the naïve proposal will be unable to explain the nature of rationality's force. We must be able to capture the fact that the requirements of rationality apply to us always, regardless of our circumstances and regardless of the contents of our attitudes. The naïve view could accommodate this central aspect of rationality only if we had persistent, invariable, decisive reason to comply with the principles of rationality. And this is simply not the case.

I am far from the first person to notice that the requirements of rationality cannot be

¹⁰ I am entirely sidestepping a number of worries one might have about our potential reasons to comply with the requirements of rationality. Most importantly, one might be worried that the reasons I am considering here are “state-given” reasons (i.e., reasons grounded in the benefits of being in the relevant state). Many current writers believe that there are no such things as state-given reasons. I readily admit that if there were any reasons for or against complying with the requirements of rationality, they would have to be state-given reasons. (I am not even sure what it would mean to have a reason for a set of attitude-types with no determinate content.)

I am ignoring these worries because even if they are legitimate, they are compatible with (and even point toward) the point I am making in this section: that the normativity of the requirements of rationality cannot be explained in terms of our reasons for compliance. If there are state-given reasons, then the arguments of this section show that they could not serve as the basis for rationality's normativity. And if there are no state-given reasons, then they could not serve as the basis for rationality's normativity because they do not exist. For helpful discussions of the debate surrounding state-given reasons, see especially Rabinowicz and Rønnow-Rasmussen 2004, Schroeder 2012, Schroeder 2013, and Appendix A of Parfit 2011a.

explained by appealing to our (alleged) reasons for having our attitudes arranged in the relevant ways.¹¹ Perhaps this is not such a surprising result; after all, if we reflect on our judgments about cases like (1), (2), and (3) from Section 1.1, it seems unlikely that our judgments are guided by the thought that the agents are failing to promote or bring about some good result. It is doubtful that we are prohibited from having our attitudes arranged in certain ways because the prohibition serves some external purpose.

If this is not a surprising conclusion, it is still a troubling one. It draws our attention to the fact that the requirements of structural rationality are rather peculiar entities. At the beginning of this section I said that rational requirements must not be brute, arbitrary ordinances on the human mind. A bit of investigation, however, raises the worry that they may be just these kinds of things. They are, taken at face value, absolute prohibitions on our combinations of epistemic and practical attitudes, and yet it is doubtful that they offer any sustained epistemic or practical benefit to anyone. Why would we be bound by any such requirements?

My goal in this dissertation is to offer a theory of structural rationality which explains what rationality is and why it binds us as it does. But before moving on to explain the process I will go through in building my proposal, I must first say more about the problems I will be addressing. In the next two sections, I will expand on the arguments of the present section and raise two serious challenges for theorizing about the requirements of rationality.¹²

¹¹ See especially Kolodny 2005, Broome 2005, and Broome 2013.

¹² Southwood 2008 makes reference to each of the problems I am exploring here.

1.4 – The Normativity Problem

In my statements of IC, BC, EI, and ME, I have said that you are “rationally required” to not have your attitudes combined in the relevant ways. As I admitted in Section 1.2, the phrase “rationally required” is awkward and unnatural. A normal person would use words like “should” or “ought” or “must” when speaking about what rationality demands of us.

I have chosen to use the phrase “rationally required” rather than “should rationally” or “rationally ought” (or some similar locution) in order to avoid a potential misunderstanding. If I were to use the words “should” or “ought,” some people might assume that I am using these terms to stand in for a reasons-based claim. It is not unusual in philosophy to say that an agent *should* X or *ought to* X when one means that the agent has *decisive reason* to X. And, as I spent the last section detailing, it is extremely doubtful that we have decisive reason to comply with the requirements of structural rationality. So if I had begun by saying that we ought not to have inconsistent intentions, and then I went on to explain that we do not have any general reasons to abstain from holding inconsistent intentions, it might have appeared that I was contradicting myself. I have chosen the phrase “rationally required” because it is less likely to create this initial confusion. I do not want it to appear that I am proposing that we have reason to hold our attitudes in any particular combinations. Were it not for this potential confusion, I would prefer to say “you ought to [...]” rather than “you are rationally required to [...]” To my ear, these phrases both express the same idea.

Avoiding this confusion is only superficially helpful, unfortunately. There is a problem here that we cannot get around just by choosing our words carefully. Whether we say “ought” or “must” or “rationally required,” the important point is that the requirements of rationality are binding or (to use a word that I have so far avoided) *normative*. To speak

metaphorically, rationality commands that we have our attitudes combined only in certain ways and these commands are legitimate—they have force over us. We need an explanation of the normativity of rationality. And we need an explanation that does not depend on the idea that we have reasons to comply with the requirements of rationality. Rationality's commands do not have force over us in virtue of our reasons for complying with those commands, because we have no such general reasons. If we take our judgments at face value, however, these commands are nevertheless normative.

Many contemporary normative philosophers would scratch their head at the thought of providing an account of rationality's normativity which is not based in reasons. These so-called *reasons fundamentalists*—most prominently Derek Parfit, Joseph Raz, and T.M. Scanlon—hold that all normative requirements are constituted by reasons.¹³ They accept that a principle is normative for an agent just in case it is one with which she ought to comply, and they claim that an agent ought to X just in case she has decisive reason to X. According to these writers, it is not just a matter of custom that we sometimes use reasons-talk and 'ought'-talk interchangeably. They think that there could not be a normative claim which was not, at root, just a claim about our reasons.

Although quite little has been written about why there could not be a non-reasons-based normative requirement, we can imagine how the argument would go. Take any claim or requirement which is alleged to be normative, but which is not based in any reasons for compliance. About any such principle, the reasons fundamentalist can ask: "If the principle

¹³ See Raz 1990 (first published in 1975) for one of the earliest works in the reasons fundamentalist tradition. Raz's view is further developed in Raz 1999, Raz 2001, and Raz 2012. Scanlon 1998 endorses a version of reasons fundamentalism, and Scanlon 2014 provides the fullest defense of the position (while also coining the term "reasons fundamentalism"). Also see Parfit 2011a and Parfit 2011b.

offers no reason for compliance, then why bother complying?” It looks as though we have only two potential routes of response. First, we could offer some consideration in virtue of which compliance is demanded. We might say: “Well, you ought to do such-and-such because...” and then go on to say something about why we are bound to comply with the principle. But if we give this kind of response, then we will have given a reason to comply with the demand, and so it will turn out that the normativity of the requirement in question is of the typical, reasons-based variety. Alternatively, we might refuse to offer considerations in favor of complying, but insist that the principle is binding in a non-reasons-based sense, relative to a non-reasons-involving standard. But if we provide this kind of response, then why should the reasons fundamentalist believe that the principle is, in fact, normative? If nothing can be said in favor of complying, then what claim can the principle make on any of us? In short, it seems that an explanation of a principle’s normativity will invariably involve an appeal to reasons, whereas a failure to explain a principle’s normativity is unsatisfactory. Call this the *Reasons Challenge*. The Reasons Challenge asks us to explain how a principle with which we have no reason to comply could bind us—how it could be normative.

Jonathan Way has recently given voice to this line of thought within the context of rationality. In investigating the requirements of structural rationality, and in confronting the possibility that we have no reason to comply with rationality’s apparent prohibitions, he raises the question of whether these requirements could still be normative. He writes:

“[T]he conclusion that there is no reason to comply with [these] requirements undermines the basic motivation for accepting such requirements. [These] requirements were introduced in order to explain what is wrong with incoherence. Since there is nothing wrong with doing what there is no reason not to do, requirements which do not entail reasons cannot serve this

purpose.”¹⁴

Way’s remarks suggest that the Reasons Challenge cannot be overcome—that a requirement whose normativity is not the normativity of reasons could not be normative at all.

One might respond to the predicament by resolving that there simply *must* be reasons to comply with the principles of rationality, even though we cannot locate them. John Broome has recently accepted such a position. Broome, whose groundbreaking work I will continue to consult throughout this project, is committed to the existence of rational requirements like those I have stated above. And he is persuaded that a requirement’s normativity must obtain in virtue of our reasons for complying with it. Having been unable to locate any general reasons to comply with rational requirements, and having admitted that he is “without [an] argument for the normativity of rationality,” Broome resolves that the requirements of rationality must nevertheless be reason-giving in a way that he cannot yet explain.¹⁵

¹⁴ See Way’s forthcoming paper, currently titled “Reasons and Rationality.”

¹⁵ Broome 2013 p. 204. Broome says that a normative requirement can be either “derivatively normative” or “non-derivatively normative.” A requirement is “derivatively normative,” in Broome’s sense, if our reasons for complying with the requirement (i.e., the reasons in virtue of which the requirement is normative) can be explained by appealing to values, or benefits, or considerations that are external to the system of requirements. The explanations I have tested (and rejected) above are all “derivative” in Broome’s sense. Because Broome cannot find any such explanation, he suggests that rationality may be non-derivatively normative: reason-giving in its own right, without appeal to any external considerations. But Broome does not say why, or how, the requirements of rationality might be reason-giving in their own right, and he says that he does “not know [an argument] that convinces [him].” Nevertheless, Broome believes that the requirements of rationality are normative in virtue of the reasons we have for complying with them. See all of Chapter 11 of Broome 2013.

In the face of such a difficult challenge, Broome's dogmatism is understandable. But it is not a legitimate solution to our problem. If we hope to offer a positive theory of rationality, then we must explain its normativity, even though we seem not to have reason to do what rationality requires.

1.5 – The Content Problem

Up until this point, I have been taking for granted that rationality is a single, identifiable (sub-)domain of principles within the broader normative domain. In the last section, I asked: in virtue of what is it true that we are required to comply with its dictates? But just as we must ask why rationality *requires* of us what it does, we must also ask why *rationality* requires of us what it does. In other words, in virtue of what do the principles within the rational domain belong to that domain, rather than to some other domain (e.g., the moral domain)? In the first section of this chapter, I noted that we use the word “irrational” to describe agents like those in (1), (2), and (3). While this labeling and grouping seems like common sense, it requires justification. A brief detour and comparison to the moral domain will help to bring out the problem I am raising.

Most contemporary moral theorists hold that there is only one ultimate moral principle, which states the single, maximally general thing that morality requires of us. For example, the Kantian holds that we ought not to murder, ought not to cheat, ought not to lie, and so on. But—to speak very loosely—these moral prohibitions are all legitimized by the single, most general moral truth: that we ought not to violate the Categorical Imperative. According to the Kantian, compliance with the Categorical Imperative involves not murdering, not cheating, not lying, and so on. While the baseline principle and the surrounding architecture is different, we find a similar structure in utilitarian theories, in

contractualist theories, and in most other systematic moral theories.

Not every moral philosopher holds such a view. Most famously, W.D. Ross offers a non-exhaustive list of distinct and irreducible *prima facie* moral duties. For Ross, there is no ultimate moral principle from which all lower-level principles, and all of our particular obligations, can be derived. Instead, Ross' investigation leads him to propose a variety of non-derivative sources of duty which interact in complicated ways to generate specific demands on our conduct.¹⁶ On Ross' account, these basic duties cannot be further unified in any interesting or informative way. Ross, sensitive to other writers' interest in a unified theory of moral requirements, writes that "loyalty to the facts is worth more than a symmetrical architectonic or a hastily reached simplicity."¹⁷ And about this Ross is surely right. Single-requirement systems have the great burden of explaining how all moral demands could have a shared source. Attempts to carry this burden often produce clumsy and unconvincing results. An ultimate requirement of morality which delivers implausible results, or which delivers plausible results only through implausible assumptions or arguments, is of little value.¹⁸

¹⁶ In fact, I believe that Ross' "*prima facie* duties" are not duties at all. They are not even what we would now refer to as *pro tanto* duties. Instead, they are *sources* of (*pro tanto*) *reasons*. So the comparison between what Ross is doing in the moral domain and what (as I will go on to explain) some writers are doing in the rational domain is somewhat inexact, because the rational requirements that I am investigating are strict, rational-duty-stating principles. This disanalogy is unimportant for the point I am making, so I will ignore it.

¹⁷ Ross 2007, p. 23.

¹⁸ Ross has no principled opposition to a unified moral theory and, indeed, he endeavors to create as unified a theory as the phenomenon allows. In response to the concern "that this catalogue

However, for those pursuing an ultimate moral requirement, concerns about symmetry, elegance, conciseness, or simplicity are, at most, secondary. The more important motivation, I believe, is that an ultimate moral requirement would allow us to explain what it takes for some demand to be a member of the moral domain. It is an important aspect of common sense that morality is unified. The liar, the cheater, and the thief all have something crucially in common; although their mistakes are different in obvious ways, their mistakes are mistakes of the same *kind*. And it is just as central to common sense that the dictates of morality are distinct in kind from the dictates of, say, etiquette. (The thief is like the liar, and neither is like the person who puts his elbows on the table during dinner.) At the same time, we are not confident that we know everything that morality demands. A successful moral theory will offer an account of the *content* of morality which explains why every moral requirement is part of the same subject matter and why it is that morality demands certain things of us rather than others. A single, ultimate requirement of morality may do just this. It is less certain that an account like Ross' can succeed in this respect.

I have been talking about the moral domain, but the surveyed considerations are just as relevant to theories of rationality. We may accept, as I do, that there are requirements concerning belief consistency, intention consistency, means-end coherence, enkrasia, and so on. And we may be inclined to see these requirements as part of the same normative subject

of the main types of duty is an unsystematic one resting on no logical principle," he writes that "[i]f further reflection discovers a perfect logical basis for this or for a better classification, so much the better" (p. 23). So, as we are reminded by David McNaughton in "An Unconnected Heap of Duties?" Ross does not hand us an arbitrary list of duties. Ross believes that some duties are basic and others derivative, and philosophical reflection leads him to conclude that the number of basic duties must be significantly larger than one. Nevertheless, many (myself included) are unsatisfied by the results of Ross' investigation.

matter. But the requirements of structural rationality cannot be a mess of principles bundled together willy-nilly. If rationality is a single, unified domain, then we need an account that explains its unity.

We saw in Section 1.4 that the failure of a naïve, reasons-based account of rational requirements has thrown into question the normativity of rationality. It is worth pointing out that in the absence of the naïve account, the content of rationality may also seem puzzling. Had the naïve view succeeded, it would have given us an explanation of what rationality is all ‘about.’ For example, suppose again that compliance with the demands of rationality is required because compliance ensures that we maximize our realized ends. In this case, it is quite obvious what unifies the domain of rationality and explains its content: rationality, as a subject matter, could be understood as a domain of prohibitive attitudinal principles which ensure the optimal realization of our ends. Of course, this is a fiction, and this is not what rationality is about. But without some account of the value or goods that rationality serves, it is deeply unclear what rationality *is* about.

Relatively little attention has been paid to the problem of explaining rationality’s content and unity, and some writers have chosen not to pursue a unified account. Again, a reference to John Broome is instructive. Broome might be considered the W.D. Ross of rationality. In fact, he may exhibit more of the Rossian quality than Ross himself.¹⁹ Broome believes that there is no hope of a unified theory and so he adopts a piecemeal approach to locating and formulating rational requirements. He proceeds simply by considering his intuitions about which mental properties an agent must have in order to be fully rational and

¹⁹ While Ross goes some length to find as much unity as possible and to reduce the number of requirements he takes as primitive, Broome has no such preoccupation.

then formulating a principle that states a demand for the properties in question. As long as the resulting formulation does not have any implausible implications, Broome is content to conclude that we have located a genuine requirement of rationality. Broome is not concerned to derive any requirement from any other, nor is he committed to there being any particular relationship between the various requirements that populate the rational domain.²⁰

We should not be satisfied with a haphazard collection of intuitively linked principles. Our bare confidence that agents like those in (1), (2), and (3) are all deserving of the title of irrationality is not philosophically adequate. If we hold that rationality is a single domain of principles, then we must be prepared to explain what holds the domain together.

Structural rationality is fascinating because although its dictates are readily accessible to common sense, it is remarkably resistant to theorization and explanation. It seems that there must be requirements of structural rationality, but we do not know where they come from, why they bind us, or what makes them what they are.

1.6 – The Arguments of this Dissertation

This dissertation is an effort to provide a theory of rationality that overcomes both the normativity problem and the content problem. The first half of the dissertation is exploratory and critical, as I attempt to both justify some of the claims made in this introductory chapter and analyze a range of competing accounts of rationality. The second half of the dissertation is positive, as I construct a novel theory that explains the unity and bindingness of rationality.

²⁰ See Broom 2013, pp. 150-151. It is striking that Broome accounts for neither the normativity nor the content of rationality.

In the next two chapters, I will investigate two very different proposals about the nature of rational requirements. In the second chapter, I will turn to Christine Korsgaard's *constitutivism* and consider its bearing on the issues raised in this introductory chapter. Korsgaard's "The Normativity of Instrumental Reason" unintentionally gave rise to the ongoing debate concerning structural rationality, and although her view about the relationship between action and normativity was not created with the goal of delivering an account of structural requirements, it is deeply informative in two respects. First, Korsgaard's work challenges an important premise I have adopted: that rational requirements constrain combinations of attitudes. Korsgaard argues that rational requirements must instead be understood as principles governing the process of thinking and reasoning, rather than as prohibitions on static sets of attitudes. Second, Korsgaard's work provides tools that could be used to respond to the normativity problem and the content problem. According to Korsgaard's constitutivism, all normative requirements (moral, prudential, rational, and so on) have a single fundamental normative source: the inescapable demand that we think and act. If we could derive the requirements of structural rationality from the constitutive standards of agency, then we may be able to explain both the normativity and the content of rationality.

My investigation of constitutivism will have mixed results. Regarding the first point, I will argue that Korsgaard and others inclined to view rationality in exclusively procedural terms are bound to miss out on crucial aspects of the phenomenon of rationality. I will use this discussion as an opportunity to reconfirm that—regardless of whether there may be some independent requirements of reasoning—there must be some static, purely structural requirements. Second, and more constructively, I will propose that we should retain Korsgaard's insight that constitutive features of our agency can generate normative demands.

Although I will contend that Korsgaard's constitutivism, as stated, does not have the resources to provide a full theory of structural rationality, I will argue that Korsgaard's work may offer the tools we need to see our way around the challenges posed by reasons fundamentalists. A Korsgaardian insight will ultimately play an important role in my account of structural rationality.

In the third chapter, I will investigate some recent work by reasons fundamentalists. The reasons fundamentalists are responsible for bringing to light the normativity problem for structural rationality explored in Section 1.4. Because there is no general reason to have particular combinations of attitudes types, the reasons fundamentalists conclude that there must be no requirements of structural rationality at all. However, they appreciate that they must attend to and make sense of our widespread, shared intuitions about rationality. Toward this end, a number of reasons fundamentalists have developed compelling *reasons-based error theories* of structural rationality. The third chapter of this dissertation is dedicated to assessing the prospects of these error theories. I will propose that even the most sophisticated version of the error theory—as provided by Niko Kolodny—is left with some uncomfortable puzzles about how to accommodate our intuitions regarding rationality and its normativity. I will argue that either the error theorist is forced to rely on at least one reasons-independent but plainly normative structural requirement (which would undermine the point of the error theory), or she will be left unable to explain core aspects of the phenomenon of rationality.

At the end of the second and third chapters, I will have reaffirmed the starting-point judgment that there are binding rational requirements on combinations of attitudes, but also reaffirmed that there is no hope for providing an account of these principles within an exclusively reasons-based system. In the fourth, fifth, and sixth chapters, I will endeavor to

construct a theory of structural rationality which resolves both the problem of content and the problem of normativity. I do so, in part, by drawing on the resources provided by the constitutivist and the reasons fundamentalist.

In the fourth chapter, I will take a brief step back from the topic of structural rationality and offer a conceptual analysis of normativity in general. I will note that the constitutivist and the reasons fundamentalist are both normative monists—they hold that there must be a single, unified normative domain, and that all normative demands must ultimately be demands of the same normative “type.” I propose that the monist thesis may be responsible for part of the mysteriousness surrounding the requirements of structural rationality. In opposition to the monist thesis, I will articulate and defend a version of *normative pluralism*, according to which there may be multiple fundamentally distinct classes of normative claims. Although I agree with the reasons fundamentalist that there must be basic normative truths about reasons, I will argue that there is nothing conceptually confused in the idea there may be something we ought to do even though we have no reason to do it. There are two steps to my argument. First, will I argue that an agent ought to comply with a requirement if and only if the agent’s violation of the requirement would be a criticizable failure or mistake. Second, will I argue that there can be a requirement the violation of which is a criticizable mistake, even if we have no reason for complying with the requirement. I will conclude that while some normative requirements invoke the ‘ought’ of reasons, others may invoke a fundamentally distinct yet wholly normative sense of ‘ought.’

Although the pluralist thesis is not itself a thesis about structural rationality (or any other substantive normative matter), I believe the thesis can be put to work in the debate at hand. I will make progress toward doing just this in the fifth chapter, where I will propose that every requirement of structural rationality can be derived from a single basic

requirement which I call the *General Requirement of Structural Rationality*. Stated concisely, the General Requirement says that we ought not to attitudinally undermine the functional success of our agential attitudes. In other words, the General Requirement prohibits holding a set of intentions and beliefs which is itself responsible for the failure of some of its members. The General Requirement resolves the content challenge by providing a road map of the entire domain of structural rationality and explaining the relationship between all of the domain's members.

Together with the resources provided in the fourth chapter, the General Requirement also allows us to resolve the normativity challenge concerning structural rationality. By understanding what structural rationality is 'about,' we gain insight into the unique sense in which we are required to comply with its demands. In the closing chapter, I will argue that we ought to comply with the General Requirement, even if it is a principle with which we have no reason to comply. Although my view is distanced from Korsgaard's, it draws on lessons provided by her work. I will propose that agents who violate the General Requirement exhibit a criticizable functional defect by undercutting the operation of their own attitudes. I will argue that because intention and belief are core and unavoidable aspects of our agency, violating the General Requirement is always a form of *agential self-undermining*. I will argue that this is enough to make true the claim that we ought to have our attitudes structured as demanded by rationality. Nevertheless, it remains true that we have no reason to be structurally rational.

2 – CONSTITUTIVISM, REASONING, AND RATIONALITY

2.1 – Assistance From Constitutivism

The apparent impropriety of harboring certain combinations of attitudes indicates that there are requirements of rationality. These requirements seem to take the form of prohibitions on our attitudinal arrangements. However, as we have seen, there are serious problems facing any attempt to provide a theory of rationality's requirements. In this chapter, I will look toward a novel account of rationality which promises to solve both the normativity problem and the content problem of rationality, but which does so by rejecting one of my central claims in Chapter 1.

Christine Korsgaard's constitutivism delivers a powerful account of rationality which promises to explain both the normativity and content of rational requirements. However, on Korsgaard's view, the requirements of rationality are not, as I have suggested, prohibitive requirements banning certain combinations of mental states. Instead, Korsgaard proposes that the requirements of rationality are principles which guide the processes of practical and theoretical reasoning. According to the Korsgaard's proposal, all normative demands, including the demands of rationality, are manifestations of a single basic normative requirement: that we comply with the constitutive standards of agency. If Korsgaard's view were successful, it would explain both what rationality is 'about' and why it binds us.

I will ultimately reject Korsgaard's proposal. But my argument in this chapter is intended to be primarily positive, rather than negative. My main ambition is not to show that the constitutivist is wrong, but instead to harvest two important results from a discussion of the constitutivist project. First, by examining an important limitation of the constitutivist proposal, I hope to secure a claim that I could make only provisionally in the opening

chapter. Specifically, I hope to show that the requirements of rationality are what I call *structural requirements*: prohibitive principles banning the simultaneous holding of certain attitudes. I will argue that the constitutivist is unable to account for these prohibitive requirements and that, as such, she cannot deliver a comprehensive account of rationality. Second, I will argue that although the constitutivist proposal cannot be right as stated, it does open the door to an important possibility. I propose that the constitutivist is right to say that features of our agency can generate legitimate normative demands. Korsgaard's arguments highlight that there is nothing conceptually confused in the idea that a requirement's normativity might be grounded in a non-reasons-based source. Later, in Chapter 4, this idea will play a crucial role in my solution to the normativity problem.

In Section 2.2, I will provide a brief overview of the general constitutivist approach to thinking about normativity and normative requirements. In Section 2.3, I will explain how constitutivism can be used to generate a theory of requirements of rationality, and I explain why this approach offers potential solutions to both the normativity problem and the content problem. In Section 2.4, I will highlight the most important way in which the constitutivist enterprise challenges my claims in Chapter 1—namely, by holding rational requirements as principles to be used by agents in reasoning. In Sections 2.5-2.7, I will argue that an account of purely structural requirements is an indispensable part of a proper theory of rationality and propose that no solely process-oriented view of rationality can deliver an explanation of structural requirements. In Section 2.8, I reflect on a crucial lesson that we may be able to take away from the constitutivist's focus on the standards of agency. Although we must reject Korsgaard's account as stated, we should retain the important insight that an agent may be bound to comply with the standards delivered by features of her agency.

2.2 – A Sketch of Korsgaard’s Constitutivism

To understand how Korsgaard’s constitutivism delivers a theory of rational requirements, we must begin by exploring what Korsgaard says about normativity. This section is dedicated to providing some background about Korsgaard’s view about normative requirements in general. Although this will take us quite far afield from the central topics of my dissertation, this detour will ultimately prove to be instructive. For the purposes of exposition, I will follow Korsgaard in focusing primarily on the case of practical norms. Toward the end of the section I will explain how epistemic matters fit within the picture. Once Korsgaard’s general account is on the table, I will move on, in the subsequent section, to explain how her view applies to the debate about the requirements of rationality.²¹

The starting point in Korsgaard’s account is the observation that we rational agents have no choice but to *act*. Korsgaard points out that not-acting is never an option for us; an effort to not-act would be futile, for it would itself involve a choice and action. If I roll over and go to sleep so as to avoid having to make a choice about what to do, then in fact I have made a choice and I have acted—I have chosen to go sleep so that I do not have to think about what else I might do. So there is no way out. As Korsgaard puts it, we rational agents are “condemned to choice and action.”²²

Korsgaard argues that this fact—the fact that we must act—has deep implications, because she believes that not just anything counts as a genuine action. Korsgaard believes that the activity of acting has constitutive standards. On her view, there are certain

²¹ My overview of Korsgaard’s position is drawn primarily from Korsgaard 2009b. For a fuller picture of Korsgaard’s view, see also Korsgaard 2009a and Korsgaard 2009c.

²² Ibid., p. 1.

conditions which must be satisfied in order for one to be legitimately engaged in acting. This means that our condemnation to action is, in fact, a condemnation to satisfying the constitutive standards of action.²³ The inescapable demand that we act carries with it a larger burden than it originally appears. According to Korsgaard, this burden is at the root of any legitimate normative requirement.

One might wonder about Korsgaard's reasons for thinking that action has constitutive standards. It is a complicated story involving multiple stages of conceptual analysis, which I will only summarize here. To begin, Korsgaard notes that when I am faced with the necessity of acting, what I am actually tasked with is "determining myself" to do something. According to Korsgaard, what separates a mere bodily movement from an action is that an action, unlike any old movement, is brought about by an agent's active choice about how to behave. I *act* only when I determine my movements.²⁴ So the demands of action are in fact demands of self-determination. Now we can ask: what would have to happen in order for me to determine myself? Korsgaard has two answers, corresponding to two closely related constitutive standards.

First, Korsgaard argues that self-determination must be a matter of actively making myself the *cause* of whatever ends I adopt. And Korsgaard says that making myself the cause

²³ This is, to my mind, the most crucial and contentious moment in the entire constitutivist argument. My primary worry is that Korsgaard slides between two uses of "act" (and "action"). It is true that we are condemned to action in the sense that, as active beings, we cannot help but do something at any given moment. But I am unsure that we are thereby bound to engage in what I have called "legitimate" action (i.e., action in a full-bodied, constitutive-standards-referencing sense). But it would be far too much for me to try to show that Korsgaard is wrong about this, and it's orthogonal to my aims in this chapter and dissertation. So I will not address my worry here.

²⁴ Ibid., p. 69.

of an end requires committing to doing whatever is required to realize that end. If I do not make myself the cause of my end in just this way, then I will not and cannot succeed in determining myself. Of course, some potential end of mine may be realized even while I lack a commitment to bringing it about, but then the end's realization will not be attributable to *me*—it will not be something that *I* have done. In that case, the end will come about *despite* me. So Korsgaard says that determining myself must include intending to bring about whatever means are crucial to the realization of my ends. In other words, Korsgaard believes that a process of self-determination—and so an action—is inherently bound by Kant's Hypothetical Imperative.

Second, Korsgaard claims that a process of self-determination must demonstrate the distinction between an agent and that agent's incentives. When I self-determine, I hold myself apart from my incentives and make a choice about what I will do and why I will do it. If my bodily movements were simply the result of various incentives and appetites going to war with each other, then those movements would not be attributable to me, because I am not identical with my incentives. When an agent *self*-determines, says Korsgaard, the agent is an entity distinguishable from the incentives she chooses to reject or endorse. Korsgaard argues that we hold ourselves distinct from our incentives only when we act on maxims that have universal form.²⁵ This part is (to say the least) a bit tricky. To see why Korsgaard thinks that this is so, consider a case in which I try to “will particularistically,” by acting on a principle which could not be willed universally. In such a case, I cannot be acting on the basis of considerations which I think could be invoked as reasons to act the same way across a range of identical or relevantly similar cases. If I acted on *that* kind of basis, then I would

²⁵ Or, at least, on maxims which we *take* to have universal form.

be willing a maxim that I took to have universal applicability. So when I will particularistically, my behavior must be dictated willy-nilly, by whatever incentives happen to have a hold on me. And when that happens I am no longer holding myself apart from my incentives and making a choice about what I will do and why I will do it—the distinction between me and my desires is lost.²⁶ And so whatever I end up doing, my behavior will not be attributable to me, and I will not have self-determined. For this reason, Korsgaard believes that a process of self-determination—and so an action—is bound by the demand that we will universalizably. In other words, self-determination is constrained by the Categorical Imperative.

On the basis of the considerations sketched in the past three paragraphs, Korsgaard concludes that there are certain standards which must be met in order for a putative process of self-determination to count as actual, legitimate self-determination. According to Korsgaard, the standards in question—the Hypothetical Imperative and Categorical Imperative—are internal to the concept of action.²⁷ This is intended to be a definitional point: Korsgaard says that “what it *means* to deliberate is to be guided by those imperatives.”²⁸ Since we agents are inescapably bound to the activity of acting, we are always bound by the Hypothetical and Categorical Imperatives.

²⁶ Ibid., p. 75.

²⁷ Korsgaard gestures to the idea that, in fact, the Categorical Imperative and Hypothetical Imperative may not be distinct principles at all. She says: “nothing counts as trying to realize some end that is not also trying to determine *yourself* to realize that end, and nothing counts as determining yourself to realize the end that is not also trying to determine your own *causality*.” Ibid., p. 80.

²⁸ Ibid., p. 131

I have been discussing practical agency, because Korsgaard's primary focus in much of her work is the practical domain. But a similar story applies to the doxastic domain. We agents are tasked not only with making choices about what to do, but also with deciding what is true (i.e., forming beliefs). And the process of determining ourselves to belief is governed by principles of logic, just as the process of determining ourselves to action is mediated by practical principles like the Hypothetical Imperative. About this, Korsgaard says:

The laws of logic govern our thoughts because if we don't follow them we just aren't thinking. Illogical thinking is not merely bad, it is *defective*, it is bad as thinking.²⁹

Korsgaard has in mind principles like *modus ponens*. She believes that such principles are constitutive of the activity of theoretical reasoning, just as principles like the Hypothetical Imperative are constitutive of the activity of practical reasoning. According to this view, if I think or reason in a way that flouts *modus ponens*, then what I am doing is not *really* thinking, just as action that violates the Hypothetical Imperative is not *really* action.

This is, admittedly, an almost comically condensed version of Korsgaard's arguments. There are many questions to ask concerning the legitimacy of the arguments just canvassed (and my exposition of those arguments). But given my purposes, I will not focus on whether these arguments succeed. Instead, I plan to assess how Korsgaard's theory contributes to the debate regarding the requirements of rationality.

2.3 – The Normativity and Content of Constitutivist Rationality

If Korsgaard's conceptual analysis concerning action and thinking is successful, then the

²⁹ Ibid., p. 32.

view delivers an account of requirements on reasoning. These requirements look like requirements of rationality, even though they are not identical to the structural requirements I put forward in Chapter 1.

Recall, for example, the principle I provisionally endorsed in the first chapter which prohibits a certain form of means-ends incoherence:

ME

You are rationally required not to [intend to X, believe that you can X only if you intend to Y, and not intend to Y].

This requirement is similar in content to Korsgaard's (and Kant's) Hypothetical Imperative, a version of which might be modeled as such:

If you intend to X and believe that you can X only if you intend to Y, then you are rationally required to intend to Y.

Now consider, as a second example, the principle I put forward in Chapter 1 which prohibits a certain form of inconsistency in belief:

BC

You are rationally required not to [believe P, believe $P \rightarrow Q$, and believe $\neg Q$].

This putative requirement of rationality has some similarities to Korsgaard's requirement concerning *modus ponens*, which says:

If you believe P and believe $P \rightarrow Q$, then you are rationally required to believe Q.

Although the structure of Korsgaard's requirements differs from the structure of the requirements that I provisionally offered in Chapter 1, it appears—at least on the surface—that Korsgaard's account may be capturing the same rational phenomena.

In a moment, I will turn my attention to assessing whether the difference between Korsgaard's requirements and my requirements is important. But before looking in that direction I must highlight a significant advantage of Korsgaard's constitutivism. Korsgaard's account promises to explain both the normativity and the content of rationality, and so

provides potential solutions to the two problems I presented in Chapter 1.

First, consider how Korsgaard's view responds to the normativity problem. According to Korsgaard's view, simply in virtue of being an agent, each of us is bound by certain requirements on our practical and theoretical reasoning. Take the example of the practical realm. If we accept Korsgaard's claim that agents are unavoidably bound to act, and Korsgaard is correct that complying with the Hypothetical Imperative is part of what it *means* to act, then agents are therefore unavoidably bound to comply with the Hypothetical Imperative. As Korsgaard puts it, "a constitutive principle for an inescapable activity is unconditionally binding."³⁰ An agent is unconditionally bound by the Hypothetical Imperative not in the sense that violating the principle is impossible, but in the sense that violating the requirement would always be a mistake on the part of the agent, given the type of being that the agent is. As a being who must always be engaged in action, a failure to engage in legitimate action is a failure of the agent.

This way of thinking about the normativity of rational requirements could be used to alleviate the pressure applied by what I have referred to as the Reasons Challenge. The Reasons Challenge asks us how a principle with which we have no reason to comply could be normative. But once we are viewing our topic through Korsgaard's lens, the request that we provide *reasons* for compliance with the requirements of rationality sounds odd. If what is being requested is some consideration that counts in favor of my complying with the Hypothetical Imperative—something like an incentive for complying, or an external benefit that I or someone else might obtain if I were to comply—Korsgaard is likely to admit that

³⁰ Ibid., p. 39.

we cannot provide an answer.³¹ But Korsgaard's view can give us a 'reason' to comply in the sense that it offers an explanation of the unavoidable defect in an agent who fails to comply. If Korsgaard's arguments are successful, then seeking an incentive to justify the demand for compliance with the Hypothetical Imperative is just wrong-headed. "Why bother complying with the Hypothetical Imperative?" asked to the Korsgaardian agent is a bit like asking a firefighter—*qua* firefighter—why he bothers to put out all of those fires. "This is just what I have to do, given the kind of being that I am," is an adequate response, and one which explains the force of the norm in question.

I will come back to Korsgaard's insight concerning normativity later in this project. But now consider how the constitutivist view may provide a solution to the content problem. According to Korsgaard's view as I have described it, the content of rationality is provided by the constitutive standards of action and thought. There is no deep puzzle on this view about what separates actual rational requirements from illegitimate requirements or about what unifies the principles that comprise the rational domain. The requirements of rationality are manifestations of the basic demand that we think and act. It is, of course, another matter to demonstrate that the constitutive standards of action and thought are able to give rise to versions of each of the requirements sketched in Chapter 1. I will not attempt here to perform any such derivations. My point is only that Korsgaard's view proposes an explanation of what unifies the requirements of rationality. Whether that explanation ultimately accommodates all of our intuitions about rationality is another matter.

I am not endorsing the view that I am discussing. Indeed, in the next section I will

³¹ One might suggest that what Korsgaard's arguments highlight is that we have some non-instrumental, purely "intrinsic" reason to comply with the requirements of rationality. Refer to fn 7 in Chapter 1; again, I must defer discussion of this question until Chapter 6.

proceed to explain why Korsgaard must be wrong in at least one crucial respect. However, it is important for my longer-term purposes to highlight how well-positioned the constitutivist is to explain the normativity and content of rationality, because the constitutivist's strategy for answering our two problems bares a striking resemblance to the strategy that I will ultimately endorse.

Toward the end of the chapter, I will come back to these issues concerning the normativity and content of rationality. But for the next several sections, my goal is to consider Korsgaard's proposal from a different perspective, in order to vindicate an aspect of the provisional proposal I offered in the first chapter.

2.4 – Rational Requirements: Narrow/Wide, Process/State

As should be obvious by this point, there is a significant difference in form between Korsgaard's requirements and the requirements I put forward in Chapter 1. Consider, again, Korsgaard's Hypothetical Imperative. It says:

If you intend to X and believe that you can X only if you intend to Y, then you are rationally required to intend to Y.

The parallel requirement that I have provisionally endorsed says:

You are rationally required not to [intend to X, believe that you can X only if you intend to Y, and not intend to Y].

Although these requirements look similar, they are importantly different. Korsgaard's requirement is a *narrow-scope* requirement, in the sense that the normative operator (“rationally required”) takes scope over only a particular attitude—the attitude which is to-be-held whenever the principle's antecedent is satisfied. By contrast, my requirements are *wide-scope* requirements, because the normative operator in each requirement takes scope over a whole set of disparate attitudes.

Because these requirements differ in their form, they also differ in their use or applicability to the agents they address. Korsgaard's requirements are what have sometimes been called *process requirements*: they are requirements which direct their subjects to undergo a certain transition in reasoning or thinking (i.e., from certain attitudes to the attitudes over which the normative operator takes scope). My requirements, on the other hand, are what have been called *state requirements*: they dictate only that their subjects not be in a particular state at any time.

It may not be immediately apparent why it matters whether we conceive of the requirements of rationality as wide-scope or narrow-scope, or as process-directed or state-directed. But it turns out that it makes a whole world of difference. Indeed, various writers, including Korsgaard, are quite conscious of the form of the requirements they endorse, and are convinced that requirements like those I put forward in Chapter 1 could not be legitimate requirements. The requirements I put forward in the first chapter were modeled as prohibitions on combinations of attitude-types in order to accommodate our basic intuitions about rationality and irrationality. But it may turn out that requirements built to match the contours of our intuitions about rationality cannot survive close scrutiny.

Numerous writers have provided this scrutiny. According to Niko Kolodny, for instance, purely structural requirements must be misstatements of the requirements of rationality, because "they do not tell you to do anything." To Kolodny's mind, wide-scope requirements barring combinations of mental states could not be counted as normative requirements at all. Kolodny says that such requirements would be mere "evaluative requirements," because they miss out on the crucial action-guiding or thought-guiding role

played by the requirements of rationality.³²

Kolodny's claim suggests that a task I have given myself in this project is an impossible one. If Kolodny is correct, then the way I formulated the requirements of rationality in the first chapter may be responsible for our inability to explain the normativity of rationality. But Kolodny's claim cannot be accepted as stated, for two reasons. First, contrary to Kolodny's suggestion, there is a clear sense in which my wide-scope requirements *do* tell you to do something. Although they do not demand that you have or form any particular attitude, they command something specific of you: that you only have your attitudes arranged in certain ways. This is something that you can do or fail to do, as it is something which is under your control (as much as anything is, anyway). Second, even if we were to agree with Kolodny that IC, BC, EI, and ME do not tell you to do anything, it is not clear why we should then accept that they are not normative. In the absence of an argument for the claim that every normative requirement is a requirement that tells you to "do something" and greater specificity about what counts as being told to "do something," there is no reason to accept that my requirements would be merely "evaluative." I see no principled reason to refuse the "normative" label to my principles. Without more being said, it strikes me as a matter of stipulation which principles we call "normative" and which we call "evaluative" and which, if any, deserve both labels.

Still, there is something important undergirding Kolodny's thought. Korsgaard has a similar worry in mind. She writes:

On my view, rational requirements do not govern combinations of our attitudes. They govern thinking, the activity of thinking; and that means that they govern someone who is actively trying to determine what she has reason to believe or to do. And thinking has a certain temporal direction. To be rational is not just to

³² Kolodny 2007, p. 372.

have a set of attitudes that happen to conform to a rational requirement. It is to follow a rational requirement, to take it as an instruction. [...] If the job of rational requirements is to govern the activities of thought and deliberation, and the point of those activities is to direct us to belief and action, then rational requirements cannot be wide scope, since wide scope requirements cannot do that job.³³

Kolodny's claim that structural requirements do not tell you to do anything is a bit obscure.

Korsgaard puts the point more clearly: there is nothing *for you to do* with a wide-scope requirement of rationality. Upon being confronted by my one of my principles in the heat of deliberation, you might question its significance to you. If you find yourself failing to intend the means you see as crucial to your ends, and I direct your attention to a requirement banning such an attitudinal state, you might respond: "Well, that's fine. But what good is your principle to me?" Korsgaard's Hypothetical Imperative, on the other hand, gives you an instruction: *take the means!*

Korsgaard's comment actually has at least two sharp points. First, she is claiming that a view like mine misses out on the role that rational requirements play in the life and thought of rational beings. As she puts it, "the claim that rational requirements are requirements on our attitudes seems to me to ignore the context in which we deploy rational requirements." There is, no doubt, something to this thought. It seems that reasoning and thinking is an activity that can go well or badly. There must be something that distinguishes proper reasoning from a jumble of thoughts bundled together, one after another, arbitrarily. And an agent engaged in reasoning does seem—consciously or not—to be appealing to certain principles to guide her in her efforts. What else could she be appealing to than principles of rationality? Second, Korsgaard's comments highlight that there may be something odd about wide-scope state requirements in the first place, since they are not the kinds of requirements

³³ Korsgaard 2009c, p. 29.

that beings like us can do anything with. Wide-scope prohibitions on combinations of attitudes may not be fit-for-purpose for agents. I suspect that Kolodny would agree with Korsgaard on both of these points. With considerations like these in mind, Korsgaard, Kolodny, and others have endorsed a view according to which the requirements of rationality must be understood as *narrow-scope* requirements on *reasoning*.

It is worth pausing to point out that, strictly speaking, there are not only two options. It is sometimes assumed that rational requirements must either take the form of narrow-scope process requirements like those endorsed by Korsgaard and Kolodny or, alternatively, as purely structural (wide-scope, state-directed) requirements like those I put forward in Chapter 1. In principle, there is at least one other view that one could defend. One could argue that there are wide-scope requirements of rationality which function as requirements on reasoning. It turns out that such a view is not conceptually incoherent, but only very bad. It is worth saying why it would be unwise to see wide-scope requirements as requirements on reasoning, because doing so will help to further clarify what underlies Korsgaard's and Kolodny's concerns.

In Chapter 1, I quickly arrived at formulations of numerous requirements of rationality by reflecting on the apparent irrationality of harboring certain combinations of attitudes. For example, from the fact that it is irrational to fail to intend to Y while both intending to X and believing that you can X only if you intend to Y, I concluded that you are rationally required not to [intend to X, believe that you can X only if you intend to Y, and not intend to Y]. And I implicitly accepted that this requirement is a static, state-based requirement. But if I were so inclined, I could argue that this principle functions as a requirement on reasoning. To defend this position, I would have to claim that you are rationally required to reason in whatever way will maintain or regain compliance with this

wide-scope requirement.³⁴ For instance, if you currently intend to X and believe that you can X only if you intend to Y, then the principle I have called “ME,” understood as a requirement on your reasoning, would direct you to do any one of the following: (I) form an intention to Y, or (II) abandon your intention to X, or (III) abandon your belief about the importance of intending to Y. Because the requirement’s normative operator is wide-scope, the requirement has no preference for how it is satisfied. Taken as a requirement on reasoning, the principle would just demand that you reason along one of the three available routes—take your pick.³⁵

This proposal is exceptionally implausible. Imagine that I intend to pass tomorrow’s exam and believe that I will pass the exam only if I plan to study all night. If I were, in Korsgaard’s terminology, to take ME as an “instruction” to guide my deliberations, then I might do so by abandoning my belief in the necessity of studying. But that is just weird. Rationality, if it governs my reasoning at all, does not do so by allowing me to reason from (a) my intention to pass the exam and my absence of an intention to study to (b) abandoning my belief that studying is crucial to passing. This is just one example. The problem is

³⁴ Kolodny was the first to note that wide-scope requirements *could* be taken as process requirements. He was also the first to point out that this would be a terrible idea. See Kolodny 2007. For a related discussion, see Broome 2007.

³⁵ It is worth noting that here and throughout this dissertation I am ignoring so-called *bootstrapping* worries about the requirements of rationality. As I suggested in Chapter 1, many writers have wondered whether we have *reason* to comply with the requirements of rationality. If we do, then it may be that we can too easily generate (i.e., “bootstrap” into existence) reasons for action and belief, especially if the requirements of rationality have narrow-scope. There are interesting questions about bootstrapping, but these are not the questions I am attending to here. See, for example, Bratman 1981, Raz 2005a, Broome 2005, and Setiya 2007.

reproducible for almost any attempted formulation of a wide-scope requirement on reasoning. In short, my wide-scope principles, if they are taken as principles governing reasoning, produce implausible results regarding what constitutes good or acceptable reasoning. If there *are* any requirements on reasoning, my requirements are not them.

So, as I said, claiming that my wide-scope requirements can function as requirements on reasoning is not a good strategy. Any plausible account of wide-scope requirements of rationality must take these requirements to be purely structural principles concerning the relationships between one's static attitudes. With this in mind, we can return to Korsgaard's concern. As Korsgaard presents the matter, wide-scope state requirements are bizarre entities, because there is nothing for us to do with them. Structural requirements seem blind to the active role that rationality plays in our agential lives.

The observations of this section leave me with a limited menu of options. Two options are the most obvious. First, in an effort to provide an account of how the requirements of rationality govern reasoning, I could abandon the provisional requirements I put forward in Chapter 1. Alternatively, I could double down on my commitment to wide-scope, purely prohibitive requirements of rationality. I favor this second option.

The next three sections are dedicated to defending my position that the requirements of rationality are wide-scope state requirements. In Section 2.5, I aim to show that wide-scope state requirements are less bizarre and useless than Korsgaard proposes and that such requirements may play an important role in our self-conception as rational beings. In Section 2.6, I will argue that these requirements are not only intelligible, but in fact an indispensable part of any comprehensive theory of rationality. Specifically, I will argue that my requirements could not, as some have proposed, be derived from an account of narrow-scope process requirements. In Section 2.7, I go even further and attempt to cast doubt on

the existence of any narrow-scope requirements of reasoning at all.

2.5 – The Role and Importance of Structural Requirements

It is not my goal to show that requirements of structural rationality can guide us in our reasoning. In Section 2.6, I will say something about the relationship between rationality and reasoning. In this section, I simply hope to illuminate what sort of role a purely state-based ban on combinations of attitudes might play in our lives. I hope to show that my wide-scope state requirements are perfectly intelligible as agent-directed requirements and, even more, that these particular requirements are central to our self-conception as rational beings.

To be sure, one role that a principle or requirement can play is to provide concrete procedural guidance to someone who is engaged in an activity. A child learning to tie her shoes can follow a principle taught to her by her parents in order to help her do the job well. Someone doing basic math in high school might appeal to the principle demanding that they deal with the brackets first in order to ensure that they arrive at the proper result. It is easy to come up with examples of requirements playing a procedural role in reasoning and decision. But this is not the only role that can be played by a principle or requirement.

Those who grew up around skating rinks should be familiar with a sign that reads, simply: “No Shoes on the Ice!” These signs issue a sort of wide-scope state requirement to their readers. The sign could be rewritten without loss of meaning to say: “You may not [stand on the ice and wear shoes].” The command offered by the sign is perfectly intelligible as an instruction, even though the guidance it offers is limited to identifying a particular no-go zone for you and your shoes. You could not use the sign’s command to make up your mind about what to do, regardless of where you are standing. For instance, if you are already complying with the sign’s demand, then you cannot appeal to the sign to help you decide

what to do next: if you are standing outside of the skating area then the sign does not ask you to enter the rink, and if you are standing on the ice in a pair of skates, the sign's order cannot help you decide whether to stay there, or to leave, or anything else. On the other hand, if you are violating the sign's directive—by standing on the ice in your shoes—the content of the command offers no recommendation about what to do to rectify the situation. The sign does not encourage you to put on some skates while staying on the ice, or to walk off the ice in your shoes, or to melt the ice from beneath your feet, or anything else. The sign says only that you may not stand on the ice in your shoes. And the sign's prohibitive demand is clear to us, and registers as a demand with which we can comply or fail to comply, even though it is of little use in deciding what we will do or how we will do it. To be clear, I am not claiming that there are no better or worse ways of complying with the sign's demand when you enter the skating arena. I am simply pointing out that the demand itself does not provide guidance concerning how you might maintain or regain compliance, and I am claiming that this does not undermine the intelligibility of the sign's dictate.

Korsgaard suggests that a view of wide-scope state requirements ignores the context in which we deploy rational requirements. But as we have just seen, a requirement is not always taken up into thought with the purpose of reasoning or making up one's mind. We may also appeal to a principle for instruction about where we may or may not go. And in some cases we reflect on a requirement just in order to assess whether we are, or someone else is, complying with its demand.

If we reflect on the circumstances in which we judge ourselves or others to be irrational, it is quite plausible that the requirements of rationality take the form of wide-scope bans on combinations of attitudes, rather than requirements to engage in some process of reasoning. When we find an agent with a set of inconsistent beliefs—your friend in (2) from

Chapter 1, for example—our inclination is to think “You shouldn't do that!” And we make this judgment without also judging that there is some particular change that she ought to make in light of this mistake. Wide-scope requirements of rationality cannot lead us from thought to thought or help us decide what to do or believe. But they can nevertheless operate as instructions that set hard limits on where our minds may go. The job of a structural requirement of rationality would not be to guide us, but rather to constrain us, by ruling out certain ways that we might put our attitudes to work.

This raises the question: but why are we subject to these constraints? Put another way: what is wrong with being in the state of having one's attitudes structured in one of the prohibited ways? Of course, these questions are simply restatements of the request that we explain the normativity and content of rationality. At this stage, I cannot offer a full explanation of why we are subject to the constraints of structural rationality. In the second half of this dissertation, I will offer a proposal. Currently, my point is only to show that there is nothing intrinsically confused or bizarre in the idea of a wide-scope, state-directed requirement.

2.6 – Can Structural Requirements Be Derived from Requirements on Reasoning?

The arguments of the previous section are not nearly enough to vindicate an account of rationality that focuses on structural requirements. Even if we accept that there is nothing bizarre in the idea of a wide-scope state requirement, and even if we find it plausible that at least some requirements of rationality take this form, we may be able to do without a theory of requirements of structural rationality. The reason is simple: we may be able to deliver a story about the requirements of structural rationality *from* an account of requirements on reasoning. About this, Korsgaard writes:

It is worth noting that if there are narrow-scope requirements, there are also wide ones. [...] You can derive a wide-scope requirement from a narrow-scope one, but you can't derive a narrow-scope requirement from a wide-scope one. So if we can explain the narrow-scope requirements in terms of the activity of reason [...] then we can explain the wide-scope ones as well.³⁶

In Section 2.4, I admitted that we cannot use a wide-scope account of rational requirements to explain or derive any requirements on reasoning. My opponent may be able to do better. Perhaps a view which sees the requirements of rationality as principles which govern reasoning could offer an explanation of why there are rational prohibitions on mental states. If this approach succeeds, then it will leave my view out in the cold, because it will account for everything in my view and a lot more. So, for the remainder of this section, I will assume that there are a variety of (not-yet-wholly-specified) requirements of procedural rationality that govern the activity of thinking and reasoning, and I will assess how far such a view can get us in explaining requirements of structural rationality.

It is easy enough to imagine how derivations would work. Suppose that Korsgaard's *modus ponens* principle is legitimate. In light of this principle, it seems obvious what is defective in an agent who simultaneously believes P, believes $P \rightarrow Q$ and believes $\neg Q$. An agent with this combination of attitudes has flouted a rational requirement on the activity of reasoning, by failing to draw the proper conclusion. So it may be that this wide-scope requirement prohibiting belief inconsistency is a mere byproduct of a requirement which governs theoretical reasoning. I said in the previous section that when we find an agent with a set of inconsistent beliefs, we are inclined to deem her irrational just in virtue of the impropriety of her attitudinal arrangement. While that may be true, our inclination to criticize her combination of mental states may be grounded in the fact that her attitudes were

³⁶ See footnote 17 of Korsgaard 2009c.

arrived at by way of a rationally impermissible process of thinking.

This argument is too quick. Certainly, there will be cases in which an agent who violates one of the wide-scope requirements of structural rationality will come into their violation state by flouting some putative requirement on reasoning. But the argument at hand proposes that *every* banned state must be one which is arrived at through a process of defective reasoning. This is not a secure assumption, for it is not necessary that every mental state is one arrived at by any form of reasoning, defective or otherwise. Non-agential processes can generate attitudes and combinations of attitudes. Beliefs and intentions often arrive without any reasoning at all, as a result of dreams, bare perceptual experiences, physical interventions into the brain, instinctual reaction, and other physiological events. There is no guarantee that a person's set of inconsistent beliefs or intentions will be the result of any process of reasoning whatsoever. An agent may awake one morning believing P , $P \rightarrow Q$, and $\neg Q$, or intending to X , intending to Y , and believing that she cannot do both. A person in this condition will be violating a requirement of structural rationality, but will not (yet, anyway) be violating any rational requirement on her reasoning.

This might sound silly. And it might seem unwise to allow my argument to hinge on fringe cases of abnormally generated attitudes. So I need to be clear about the point I am trying to make. Although I suspect that there are, in fact, many real-life cases of structural irrationality that are not the result of any processes of thinking at all, it is not important to my argument that such cases occur often (or ever). I am making a point about whether a requirement on mental states can be strictly speaking *derived* from a requirement on processes of thought. I am saying that it cannot, because states are not (as a logical or metaphysical matter) necessarily the product of processes. Consider, again, my analogy regarding the signs posted at skating rinks that demand that we not stand on the ice in our shoes. Could we

derive this prohibition from a system of procedural requirements governing our actions and movements with respect to the ice rink? Strictly speaking, the answer is *No*. You could come to be standing on the ice in your shoes without having done anything at all. You could be spontaneously generated on the ice in your shoes or, alternatively, someone else might knock you unconscious on the street, drag you onto the ice, and wake you up. If all we have at our disposal are requirements governing how people may choose to move with respect to the ice rink, then we will not be able to prohibit people from standing on the ice in their shoes *per se*. The same is true in the case of mental states. Requirements on reasoning cannot be used to derive strict bans on combinations of mental states, because mental states need not be the result of any processes of reasoning or thought.

An opponent might say that she can accept this result. With her focus set on reasoning alone, my opponent could claim that there is not necessarily any irrationality in a person whose mental states have not arisen from defective reasoning. The proponent of process requirements may admit that she cannot derive principles like IC, BC, ME, and so on, but may also claim that she is willing to reject these requirements altogether.

It would strain credulity to observe an agent with a set of inconsistent intentions and claim that unless he made a mistake in his reasoning he is making no mistake at all and is unbound by any requirement. So my opponent will need *something* to say about this kind of agent. She might argue, then, that an agent with an intuitively unacceptable combination of mental states would be irrational only if he did not go on to reason his way out of the relevant state in some rationally permissible fashion. For example, if an agent were to awake from a science experiment in a state of intending to X, intending to Y, and believing that he could not both X and Y, my opponent could argue that the agent would be irrational only if he did not exit that state by reasoning in accordance with some narrow-scope principle of

practical reasoning.

But if an agent with a set of inconsistent or incoherent attitudes is always required to reason out of the state that she is in, we need an explanation of why this is so. But it will not be easy to provide a general account of why one is always rationally required to reason out of an intuitively improper attitudinal state that does not rely on or take for granted the rational impropriety of *being in the relevant state*. For example, as I have argued above, IC cannot be derived directly from any requirements on reasoning. My opponent may be happy to reject IC. My opponent will then have to explain why, as a general matter, any agent who intends to X, intends to Y, and believes that she cannot both X and Y will be irrational if she does not reason her way out of her attitudinal circumstances. But the explanation cannot make reference to the irrationality or defectiveness of having the relevant combination of attitudes, because doing so would grant me too much. If my opponent explains the rational requirement to reason out of some state by pointing out the impropriety of being in that state, then that very explanation could be used to explain a requirement of structural rationality instead.

In fact, this is a concern for any attempt to justify requirements on reasoning more generally. When we ask what justifies some particular requirement on our reasoning, it is difficult to offer an explanation that does not make reference to the bad outcomes that a failure of compliance with the principle would produce. But those bad outcomes—the states that I would like to count as irrational—are what we are now attempting to explain by appealing to the requirements on reasoning. Of course, without a full list of the requirements on reasoning that a proponent of the narrow-scope view of rational requirements is offering, I cannot conclusively prove my point here. However, I believe that the burden of proof rests with my opponent.

Even if we put aside what I have said so far and imagine that we can, somehow, provide a system of requirements on reasoning that can be used to derive the requirements of structural rationality, I am skeptical that such an account would be explanatorily satisfactory. A derivation of the requirements of structural rationality from some principles on reasoning would, by its nature, suggest that structural requirements are parasitic on the principles of reasoning. To my mind, this is an unnatural and counter-intuitive form of explanation, even if it is extensionally adequate. When I introduced the requirements of rationality in the opening section of this project, I made no reference to reasoning into (or out of) any state—I did nothing more than characterize the combination of mental states harbored by the agents in question. And that is enough, I think, to generate our judgments about the irrationality of the states that such agents are in.

At this stage, I have not claimed that there are no rational requirements that govern reasoning. I have said only that we cannot do without the requirements of structural rationality and that we cannot derive these important prohibitive requirements from requirements on reasoning. If there are requirements on reasoning, they must be independent of the wide-scope state requirements of rationality.

2.7 – Skepticism About Requirements on Reasoning

The previous two sections, taken together, suffice as a response to part of the challenge that I am considering. What I have tried to show is that structural requirements are intelligible and important aspects of a proper theory of rationality. But I have not said anything to suggest that they could be the whole story about rational requirements. Even if there is important work that can only be done by wide-scope prohibitive principles, it may be that there is important work that can only be done by procedural requirements as well. This

might be an appropriate ending point to this line of inquiry. Although philosophers have been keen to try to make us choose *between* a theory of structural requirements and a theory of requirements on reasoning, there could be both kinds of requirements. Perhaps the domain of rationality is just a bit messy.

In fact, I suspect that things are not this messy. Although I am not prepared to rule out the possibility that there are requirements on reasoning, I doubt that any such requirements exist. This section is dedicated to expressing my skepticism about narrow-scope, process-directed requirements of rationality. To begin, I will attend to some candidate requirements on reasoning and explain why I am not satisfied with them. This will eventually lead me to offer a more general explanation about the relationship between rationality and reasoning.

If there is such a thing as a rational requirement that guides the activity of theoretical reasoning, you might expect that it has something to do with *modus ponens*. Consider an example of how an agent might be required to employ modus ponens in his reasoning. Imagine that Brett hears a jingle outside of his front door. It is a very familiar sound which he has heard countless times before. Immediately, he comes to believe that what he is hearing is the jingling of keys just outside of his front door. Given how this has gone many times in the past, he is already carrying a background belief that when he hears the jingling of keys just outside his front door, someone will unlock the door and walk inside. If Brett's reasoning is governed by a narrow-scope requirement based on modus ponens—if when he believes P and believes $P \rightarrow Q$, he is rationally required to then believe Q —then, rationally, he must now form the belief that someone will unlock the door and walk inside. This might sound pretty plausible. If Brett did not draw the conclusion and arrive at the new belief, he might appear to be exhibiting a breakdown in his reasoning.

But whether it seems plausible that Brett is required to draw the conclusion via modus ponens actually depends on what *else* Brett believes. Suppose that at the same time that Brett forms his first two beliefs, he recalls that neurotoxins eradicated the rest of humanity last week. Upon having this recollection, Brett judges that it is overwhelmingly unlikely that someone will unlock the door and enter his home and, so, at once he believes that it is false that there is anyone standing outside and about to come in. In light of this, he also loses his initial belief about hearing the jingling of keys just outside of his front door. “It must just be the wind chime,” he thinks.

Brett has not reasoned in accordance with modus ponens from his initial two beliefs. If he were rationally required to do so, then it would be a mistake in reasoning to not draw the conclusion that someone standing outside is about to enter. However, intuitively, he makes no mistake in not reasoning to this conclusion. It seems that the way he has reasoned is rationally immaculate. Examples like this are easily reproducible. There cannot be a requirement of rationality which says that we must always reason in accordance with modus ponens. To be clear, I am not here denying that Brett is reasoning in accordance with some legitimate principle. I am pointing out only that his reasoning is not subject to a requirement based on modus ponens.

The same difficulty emerges when we try to construct a requirement which governs the process of practical reasoning. Recall, again, the version of the Hypothetical Imperative which says:

If you intend to X and believe that you can X only if you intend to Y, then you are rationally required to intend to Y.

It is very unlikely that there is any such requirement on our reasoning. Suppose that you extensively plan to buy a new boat tomorrow morning—picking out the model, color, the dealership, the time of purchase, and so on. Now suppose that at some point this evening

you come to believe that robbing a bank tonight is the indispensable means to your successfully buying a boat in the morning. (Until after dinner, you had forgotten that you do not have nearly enough money for a boat.) The principle stated above would direct you to reason to an intention to rob a bank. Presumably, you will not in fact go ahead and intend to rob a bank. For the sake of ease, we can imagine that you already have a background commitment to abstain from committing crimes. In light of this preexisting commitment, rather than forming an intention to rob a bank, you abandon your intention to buy a boat tomorrow morning. This seems to be a rationally acceptable process of reasoning. As stated, however, your reasoning violates the Imperative. This suggests that there is no such requirement on practical reasoning.

This is not enough to show that there are no rational requirements on reasoning. All that I have argued so far is that some of the obvious candidate principles are illegitimate. It may be that there are requirements on reasoning which explain what is going on in both of these cases. Here is one potential explanation. It could be argued that what makes the agents' reasoning rationally acceptable is their compliance with their own normative judgments. In the theoretical case, Brett has a normatively loaded belief regarding his circumstances: he believes that the evidence (the neurotoxin apocalypse) overwhelmingly favors the proposition that there is no one standing outside and about to come in. When he does not draw a conclusion from his other two beliefs—the conclusion that someone will unlock the door and walk inside—this is because he has responded to his normative judgment in a certain way. In the practical case, when you abandon your intention to buy a boat in the morning, rather than intending to rob a bank, you do so in accordance with your judgment that you have strong reason not to rob a bank. In both cases, then, one might argue that the agent's rationality in reasoning is secured by the fact that there is an appropriate alignment between

the agent's normative judgments (about what to do or believe) and the agent's other attitudes. Moreover, it could be argued that the agents would be irrational if they did not exhibit this alignment between their normative judgments and their other attitudes. If, for example, Brett reasons from P and $P \rightarrow Q$ to Q , but does so while believing that there is insufficient evidence in support of Q (or, even more, conclusive evidence against Q , given humanity's destruction), then he may be making a rational mistake.

Considerations like these are no doubt behind a proposal put forward by Joseph Raz and Niko Kolodny which says that rationality, as a whole, is rooted in the demand that we respond appropriately to our perceived reasons. Inspired by Raz's work, Kolodny put forward the following two principles of rationality:³⁷

(C+) If one believes that one has decisive reason to have attitude A , then one is rationally required to have attitude A .

(C-) If one believes that one has insufficient reason to have attitude A , then one is rationally required not to have attitude A .

Kolodny proposes that any other apparent procedural requirements on our reasoning are derivable from these two basic requirements. This approach promises to explain why it is that we are often rationally required to, for example, reason in accordance with modus ponens. In many cases, an agent who believes P and believes $P \rightarrow Q$ will thereby be in the position of taking himself to have conclusive reason to believe Q . If the agent is in this position, then he will, by Kolodny's first principle, be rationally required to believe Q . So the agent will in this case be rationally required to reason in accordance with modus ponens, even though there is no general rational requirement demanding that one reason in

³⁷ See Kolodny 2005, pp. 541-542 and p. 557. I have superficially changed the presentation of Kolodny's principles, simply to bring out their narrow-scope structure.

accordance with modus ponens (or any other principle of logic, for that matter). Perhaps we should follow Kolodny and identify the enkratic requirements as the basis of rationality's imposition on our reasoning.

I share at least some of the judgments behind the Raz-Kolodny proposal. Like Raz and Kolodny, I am confident that we are not, as general matter, rationally required to reason in accordance with any set principles of logic. And I also agree that an agent who believes that he has conclusive reason to have an attitude but does not have that attitude is irrational. But I cannot endorse Kolodny's proposal as a proposal about requirements on reasoning, because I am skeptical that the principles listed directly above are requirements *on reasoning*. If what we are looking for are requirements that could govern processes of thinking or reasoning or attitudinal adjustment, then these requirements seem either trivial or useless.

On the proposal at hand, we are to imagine that there is some process of reasoning or thought that can bring an agent from a normative judgment to the attitude that the judgment represents as to-be-held, and that one must engage in this activity in order to satisfy a requirement of procedural rationality. When an agent inspects the evidence before her and forms a judgment like *I have decisive reason to believe P*, then, if she is rational, she will transition in thought from that normative judgment and come to believe P. This strikes me as an awkward and unconvincing description of what does or should happen. We do not reason or think our way *from* normative judgments like these *to* the attitudes that our judgment represents as to-be-held. When we have reached a normative judgment like *I have decisive reason to believe P*, our reasoning and thinking is finished. There need not and should not be any temporal, attitudinal, or procedural gap between the formation of the normative verdict and the formation of the attitude that is judged as to-be-held. There need not be any process of moving from one to the other—the formation of the normative belief brings

about, at the same time, the attitude that is deemed as to-be-held. Certainly, we have no use for any principle that would “guide” us in some process like this.

I am not denying that an agent can have her attitudes out of alignment with her normative judgments. An agent may, for example, believe that she has decisive reason to intend to X, but not intend to X. Such an agent is suffering from a form of akrasia in intention. Indeed, in example (3) at the outset of Chapter 1, I am exhibiting precisely this form of irrationality when I fail to intend to book a flight to visit my long-lost relatives. I have not changed my opinion about the irrationality of agents in such circumstances. What I am saying is that a breakdown between one’s normative judgments and one’s other attitudes is not easily understood as a procedural failing, as a breakdown in thinking or reasoning.³⁸ It is much more plausible that enkratic requirements are wide-scope state requirements. So even these enkratic principles, which stood the best chance of surviving scrutiny, are unlikely candidates as legitimate requirements of procedural rationality.

To be clear: I am not yet rejecting the possibility that we might use enkratic principles to explain or derive other principles of rationality. I will attend to that possibility in Chapter 3. For the moment, I am only registering my skepticism about grounding a network of principles of *procedural* rationality in requirements of (procedural) enkrasia.

The arguments of this section are suggestive rather than definitive. I cannot here prove that there are no requirements of procedural rationality. And it is not important to the positive proposal that I will go on to offer that there are no procedural requirements of rationality. But I hope that what I have argued indicates where the preponderance of the evidence currently lies. Earlier in this chapter, I encountered a position which tells us that the

³⁸ I will come back to this point in Chapter 5, when I discuss how we should conceive of structural requirements governing enkrasia.

only legitimate requirements of rationality are narrow-scope procedural requirements, and according to which my purely structural requirements, if intelligible at all, can be explained away as mere byproducts of the principles which govern reasoning or thinking. I have tried to show that, quite differently, not only is it impossible to derive my wide-scope requirements from requirements governing reasoning, but that it is quite difficult to find intuitively plausible requirements on reasoning at all.

Before closing this section, I must clear up two potential misunderstandings that could arise from what I have argued so far.

First, it may seem that my arguments in this section conflict with my earlier claims about the requirements of structural rationality. For example, in this section, I have argued that there is no rational requirement to form an intention to Y whenever you intend to X and believe that you can X only if you intend to Y. I gave the example of you abandoning your intention to buy a boat after coming to believe that you could buy a boat only if you intended to rob a bank. In this example, as described, there is moment at which you both intend to X and believe that you can X only if you Y, but during which you do not intend to Y. During this moment, you are violating the principle that I have labeled “ME.” But given that I have said that your reasoning is unassailable in the case, it may seem that I am committed to saying that you are at no point irrational. This would be odd, as I have endorsed ME as a legitimate requirement of structural rationality.

However, I did not say that that you are at no point irrational in the example. I have argued only that you are not irrational in virtue of your reasoning. Specifically, your irrationality is not due to your *failing to form an intention to Y*. Indeed, I believe that you are irrational in this case, and the proposal I will put forward in Chapter 5 will explain what your mistake consists in. For the moment, my point is only that your reasoning, *qua* reasoning, is

unassailable.

Second, my skepticism about rational requirements on reasoning may make it appear that I have nothing at all to say about what is going on when we reason. If there are not any rational requirements governing our reasoning, then is reasoning just a mishmash of disconnected thoughts? Am I denying that we generally reason in accordance with the rules of logic? To both of these questions: *No*. Reasoning is what we do when we are trying to reach judgments about how we should act and what we should believe. When we are engaged in this process, we routinely put to use the rules of logic, just like we call on the rules of mathematics and even the laws of nature (as we understand them). Facts about the way that the world is organized (logically, mathematically, physically, and so on) are of considerable use in this effort. I am not skeptical about any of this. I am skeptical only that there are requirements of rationality demanding that we reason in accordance with the rules of logic, or with any other fixed rules.

2.8 – Lessons From Constitutivism

I began this chapter by talking about Korsgaard's constitutivism. But I have since travelled a long way from that starting point. In closing, I must restate the two important conclusions that I have tried to harvest from this discussion.

The first major result, which has been the subject of the second half of this chapter, is a vindication of the intuition that there are purely structural requirements of rationality. In looking at the limitations on Korsgaard's exclusively procedural conception of rationality, I have tried to show both that it faces its own set of obstacles and that it could not be used to derive a system of wide-scope state requirements of rationality. And, at the same time, I have argued that purely structural requirements are intelligible normative principles which play an

important role in our thinking about rationality and irrationality.

While disagreeing with Korsgaard on these important issues, I have suggested that there is a crucial insight that we might carry forward. In the first half of the chapter, I explored Korsgaard's account of the foundations of normativity. Korsgaard argues that a requirement could have force over an agent if it were derived from the constitutive standards of agency. The underlying idea is that if an agent's non-compliance with a putative requirement would unavoidably involve an agential failure, then the agent is bound, as the type of being that she is, to comply with the requirement in question. Though I am not optimistic about Korsgaard's arguments concerning the constitutive standards of agency, I will put her general insight to work in my own proposal. I will come back to Korsgaard's arguments in Chapters 4 and 6.

3 – REASONS FUNDAMENTALIST ERROR THEORIES

3.1 – The Return of the Reasons Fundamentalists

Although there seem to be requirements of structural rationality—wide-scope prohibitions on our combinations of attitudes—we have little idea where they come from, what binds them together as members of a single domain of principles, or in virtue of what we are required to comply with their dictates. In Chapter 4, I will begin building an account of rationality which is intended to answer these questions.

Before moving on to the constructive half of this project, however, I will first take a deeper look at the reasons-based views that got us into this trouble in the first place. In Chapter 1, we saw that the requirements of structural rationality cannot be explained by appealing to our reasons for having our attitudes arranged only in certain ways. But some contemporary normative theorists—the reasons fundamentalists—propose that any normative requirement must be normative simply in virtue of the reasons we have for complying with it. Given the fact that we have no general reasons to comply with the requirements of rationality, these reasons fundamentalists deny that there are normative requirements of structural rationality at all.

While the reasons fundamentalists are free to deny the existence of the requirements of rationality, this denial comes at a cost. They cannot completely ignore the starting-point judgments that put this project, and many others, in motion. If the reasons fundamentalists wish to reject the requirements of structural rationality, they must also offer an explanation of why we possess these (allegedly mistaken) judgments about rationality. Unsurprisingly, a number of reasons fundamentalists have done precisely this, by constructing reasons-based error theories of structural rationality. This chapter is dedicated to exploring the plausibility

of these error-theoretic efforts.

In Sections 3.2 and 3.3, I will sketch and discuss two related versions of the reasons-based error theory: the original error theory offered by Joseph Raz, and the new error theory developed by Niko Kolodny. In Sections 3.4, I will construct an example that I will use to illustrate a network of problems facing the error theories. In Sections 3.5 and 3.6, I explore those problems and a number of potential error-theoretic responses. Ultimately, I will argue that the reasons-based error theorist is left with a choice between (a) leaving unexplained some of our most basic intuitions about rationality and (b) abandoning the error theory by accepting at least one reasons-independent normative requirement.

3.2 – The Early Razian Error Theory

In a pair of seminal articles on instrumental rationality, Joseph Raz proposes that there are not any normative requirements of rationality at all. Raz, one of the earliest proponents of reasons fundamentalism, does not deny the intuitive appearance of requirements of rationality. But he believes that we can make sense of our judgments about rationality without accepting that there are any genuine, normative requirements. Raz's illuminating (though sometimes cryptic) work set the agenda for constructing a reasons-based error theory of structural rationality.

The crux of Raz's view comes in the following passage:

The fallacy to avoid is the thought that irrationality [...] occurs only if one fails to conform to a reason [...]. It can consist in faulty functioning, that is in ways of thinking and of forming beliefs or intentions, and so on, which do not conform to standards of rationality.³⁹

³⁹ Raz (2005a), p. 15.

One might wonder what Raz means by “faulty functioning,” as there are many ways in which an agent might be said to function poorly. But Raz has in mind a very specific form of faulty functioning. Speaking to the issue of so-called instrumental irrationality (which I have been referring to primarily as “means-end incoherence”), Raz says that irrationality consists in a “malfunctioning of our capacity to react properly to perceived reasons that manifests itself in failure to pursue [the believed] available means to our ends.”⁴⁰ In other words, Raz tells us that instrumental irrationality is simply a failure of *enkrasia*.

Although Raz’s comments are directed to the phenomenon of means-end incoherence, some writers have understood Raz to be offering a fully general proposal about rationality. And regardless of Raz’s intentions, the proposal certainly can be generalized. Inspired by Raz’s work, Niko Kolodny has offered one such explicitly generalized account. In “Why Be Rational?” Kolodny proposes that every putative requirement of rationality could be derived from a pair of “core” requirements, each demanding that we respond appropriately to our perceived reasons, by having every attitude we take ourselves to have decisive reason to hold and by having no attitude we see ourselves as having insufficient reason to hold.⁴¹ If Kolodny is right, then all of rationality is rooted in the requirements of *enkrasia*.

I first introduced this proposal in Chapter 2, within the context of assessing whether there might be some requirements of procedural rationality. I have argued that we should not see the requirements of *enkrasia* as procedural requirements, but instead as wide-scope requirements of structural rationality. For the purposes of this chapter, that debate is largely

⁴⁰ Ibid.

⁴¹ Scanlon offers a similar proposal in Scanlon 2007.

irrelevant. And, in fact, as we will see, Kolodny's new version of the error theory is specifically focused on explaining away the requirements of structural rationality, rather than providing an account of procedural rationality. So I will say nothing more about the state/process debate.

As I indicated at the beginning of this section, these Razian views about rationality are *error theories*. What I mean by this is that these accounts propose that there are no legitimate (i.e., normative) requirements of rationality at all, while at the same time seeking to make sense of and explain away our preexisting judgments about rationality. It may not be immediately obvious why these views should count as error theories, given that they appear to put forward a positive account concerning the foundations of rationality (in enkrasia).

Raz and Kolodny do not believe that there are any normative rational requirements, including the requirements of enkrasia. While Raz and Kolodny have the ambition of showing that every apparent rational failure is an enkratic failure, they also hope to show that enkratic failures are themselves but *apparent* failures, rather than genuine normative failures. How we are to understand the merely apparent normativity of enkrasia is a topic that I will save until Section 3.6. For now, I am merely flagging this important aspect of the error theory.

Setting aside questions about rationality's normativity, I will for the moment focus on rationality's content. In order for the Razian error theory to succeed, it must successfully account for all of our central intuitions about rationality and irrationality. And in order to do this, it must show that every standard requirement of structural rationality—those like IC, BC, and ME—can be derived from the demands of enkrasia. As it so happens, this is an extremely challenging problem for the error theory.

Some cases of structural irrationality do seem to involve enkratic failure. Suppose,

for example, that you intend to make dinner for your partner who will be returning home from abroad this evening. You know that you will succeed in making the meal only if you get off the couch right now and head to the grocery store immediately. But without abandoning your intention to make dinner, you have and form no intention to get off the couch to go to the store, instead opting to watch another episode of your favorite television show. At first blush, we would be inclined to appeal to ME to explain your irrationality. But the Razian error theorist will insist that your failure is in fact an enkratic failure: you are failing to have the attitudes deemed appropriate by your perceived reasons. In the case at hand, this is quite plausible. Perhaps you intend to make dinner for your spouse because you view yourself as having decisive reason to do so. Accordingly, we can imagine, you view yourself as having decisive reasons to take the necessary means (i.e., get off the couch and go to the store) in order to realize your end (i.e., making the meal). So when you lack an intention to get off the couch, you are lacking an attitude which you see as demanded by your reasons. As such, your means-end incoherence can be categorized as an enkratic failure. We do not need to appeal to ME in order to explain what is going wrong with you, as we can appeal directly to the demands of enkrasia.

While it is easy to imagine cases which are accommodated by the error theory, it is deeply implausible that the requirements of enkrasia can be used to completely discard every other requirement of structural rationality. Cases of means-end incoherence need not exhibit the simple structure displayed in the example of the last paragraph. For example, an agent may intend to realize an end for which she sees herself as having merely sufficient reason. In such a case, the agent may also view herself as having merely sufficient reason to pursue the necessary means. As such, if she fails to intend the (believed) necessary means, she will merely be lacking an intention for which she views herself as having sufficient reason. But

enkrasia does not demand that we hold every attitude for which we see ourselves as having sufficient reason. So while the agent in question will be means-end incoherent, we will not be able to explain what is wrong with her by appealing only to enkrasia. While some cases of instrumental irrationality involve enkratic failure, some may not. This means that we cannot so quickly discard ME.

Other forms of structural irrationality are also difficult for the Razian error theory to explain. Kolodny admits as much at the end of “Why Be Rational?” He writes:

[T]he requirement to intend the apparent means if one intends the end, does not seem to have the form of either of the core requirements. [...] Likewise, rational requirements to have logically consistent beliefs do not seem reducible to the core requirements. After all, one is rationally required to hold logically consistent beliefs, no matter what their subject matter: that is, whether or not they are beliefs about reasons for attitudes.⁴²

Despite this worrisome acknowledgment, Kolodny ends his paper on a hopeful note: he suggests that, despite appearances, every requirement of rationality may be derivable from his core requirements. But no such comprehensive derivation scheme has been forthcoming, and I am inclined to believe that the prospects are bleak.⁴³

⁴² Kolodny 2005, pp. 559-560.

⁴³ In this chapter, and in the rest of this project, I will be ignoring a range of views which have a great deal in common with the reasons-based error theories of Raz and Kolodny. According to these other views, which we might call *subjective-reasons theories of rationality*, rationality is (in whole or in part) a matter of responding appropriately to our “subjective reasons.” Views of this form have been developed in Schroeder 2009 and Parfit 2011a.

At first glance, a subjective-reasons view may look identical to the early Razian error theory, because to say that (A) *one is rationality required to respond appropriately to one’s perceived reasons* may seem to amount to the same thing as saying that (B) *one is rationally required to do what one’s subjective reasons require*. But there may be an important difference, depending on how we analyze the notion of a subjective reason. If we say that one has subjective reason to X just in case one believes that one has

3.3 – Kolodny’s New Error Theory

The failure of the basic Razian account is not the end of the road for the reasons fundamentalist error theories. More recently, Kolodny has gone to great lengths to provide an error theory with more explanatory power than the old Razian account.⁴⁴ Though

reason to X, then the subjective-reasons account may be identical to the Razian error theory. But the notion of a subjective reason could be defined more broadly. We might say, for example, that one has subjective reason to X just in case one has beliefs whose truth would make it that case that one (in fact) has reason to X. This would allow one might have subjective reason to X while having no beliefs at all about one’s reasons for X-ing. (For example, one might believe that it is raining outside, even though it is not raining outside. In such a case, on this broad account of subjective reasons, one would have a subjective reason to take an umbrella when leaving the house, even if one has no beliefs about having a reason to take an umbrella, simply because one would have a reason to take an umbrella if one’s belief about the rain were true.) How much overlap there is between the enkrasia-grounded Razian error theory and the subjective-reasons theory depends on how the notion of a subjective reason is unpacked.

I am ignoring these subjective-reasons theories of rationality because they are less plausible and less well-developed than the error theories I am discussing in this chapter. In Parfit 2011a, Parfit is agnostic about whether the entirety the rational domain can be explained by a subjective-reasons view or whether there may be completely independent requirements of rationality. See p. 36 and pp. 118-125 of Parfit 2011a. Schroeder has developed his account only with respect to means-end coherence, and it is doubtful that it could be extended to account for every requirement of rationality. (Although, the introduction to Schroeder’s forthcoming paper entitled “What Makes Reasons Sufficient?” suggests that Schroeder may hold such a view.) Additionally, the arguments I press against Raz and Kolodny in this chapter are equally applicable to subjective-reasons theories.

One might ask whether subjective-reasons theories are error theories. This is an interesting issue and one that I cannot settle here. The short answer is that it is possible to be a subjective-reasons theorist who is a realist about rational requirements and also possible to be a subjective-reasons theorist who is an error theorist about rational requirements.

⁴⁴ See especially Kolodny 2008a and Kolodny 2008b. In this section, I am aiming to condense Kolodny’s extremely complicated arguments into a more concise and palatable form.

Kolodny's arguments are richer and more complex than I can do justice to in a short space, I will try to provide a sketch that brings out its most salient details.

Kolodny's new error theory does not attempt to show that every failure of structural rationality involves an enkratic failure. In fact, the core of the new error theory has nothing to do with responsiveness to perceived reasons. Instead, Kolodny argues that whenever a person is in a position that we normally associate with the violation of a structural requirement, they are in fact failing to respond appropriately to some of their actual reasons. At first glance, this is bound to look like a version of the naïve reasons-based view rejected in Chapter 1. But Kolodny's suggestion is not that we have decisive reasons to hold our attitudes only in certain kinds of combinations. Instead, Kolodny proposes that when people have certain combinations of attitudes (i.e., those we normally associate with structural irrationality), they are in a position which guarantees that they are failing to comply with some demand of their reasons which itself has nothing to do with attitude arrangement.

Some care is required to explain why Kolodny's new view is distinct from the naïve view. According to the naïve reasons-based view rejected in Chapter 1, structural irrationality is a matter of failing to respond to our wide-scope reasons to have (or not have) a particular arrangement of attitudes. On the naïve view, there are normative principles which say, roughly, that *you have decisive reason not to [attitude A, attitude B, attitude C ...]*. This naïve view is a realist view about structural rationality, because it proposes that there are legitimate, wide-scope, reasons-based requirements on our combinations of attitudes. According to Kolodny's new error theory, there are no such wide-scope requirements. Instead, Kolodny proposes that instances of apparent structural irrationality can be explained away by appealing to two basic demands of our reasons: to only hold attitudes for which we have sufficient reason and to hold every attitude for which we have decisive reason. Kolodny

argues that agents we would normally interpret as structurally irrational are in fact having a non-structural problem, as there is bound to be some specific spot within their set of attitudes where they are going wrong with respect to their reasons.

Kolodny's view is an error theory because it denies that there are any structural requirements *per se*—there are, on this view, only the typical demands of our reasons, and it just so happens that agents with certain combinations of attitudes are bound to fail with respect to their reasons. If Kolodny is correct, then I moved too quickly in Chapter 1 when I inferred from cases like (1), (2), and (3) that there are wide-scope state requirements of rationality. The new error theory tells us that what we should take away from those examples is that such agents are bound to be making a mistake somewhere within their attitudes, rather than that their mistake resides in the combination of attitudes as such.

To make Kolodny's new theory more tractable, consider a familiar putative requirement of rationality: the requirement of consistency in intention (IC). I have proposed that there is a requirement of rationality demanding that you not [intend to X, intend to Y, and believe that you cannot both X and Y]. If the error theory is correct, then whenever you have the relevant combination of attitudes, you will have gone wrong with respect to your reasons for some particular attitude or attitudes within the set. Kolodny considers a number of potential explanations concerning where you might have gone wrong. For expository purposes, I will consider only one very straightforward potential explanation.

Suppose that whenever one intends to X, one also thereby believes that one will X—that an intention comes bundled with a belief about its own success.⁴⁵ This is a contentious

⁴⁵ This claim about intention has been defended, in various forms, in Harman 1976, Wallace 2001, Setiya 2007, and elsewhere. It is worth noting that these writers have tried to use this claim to ground requirements of practical rationality in requirements of epistemic rationality. In short, so-

assumption, but for present purposes we can suppose that it is true. If this is so, then whenever one intends to X, intends to Y, and believes that one cannot both X and Y, one thereby believes that one will X, believes that one will Y, and believes that one cannot both X and Y. These three beliefs have logically incompatible contents. Now (an additional assumption) suppose that it is not possible to jointly have sufficient reason to believe P, Q, and $(P \rightarrow \neg Q)$. This is a plausible claim, since evidence in favor of P is, on its face, evidence against Q, if $(P \rightarrow \neg Q)$ is true. If this is all correct, then it is clear why inconsistency in intention involves a failure of compliance with your reasons: if you intend to X, intend to Y, and believe that you cannot both X and Y, then you will have at least one belief for which you lack sufficient reason. The normative failure of the agent who suffers from inconsistency in intention is simply that she has at least one belief that is unsupported by her reasons. In this explanation, the putative structural requirement (IC) becomes unnecessary baggage that we can jettison. And, of course, the same explanation could be used for requirements concerning consistency in belief (such as BC).

I provide this sample explanation only to illustrate the mechanics of the new error theory. I am not supposing that the explanation just canvassed makes the most persuasive case possible. This explanation relies on controversial theses about the epistemic concomitants of intention and the possible relationships between epistemic reasons.

Kolodny realizes that the argument is contentious. In light of this, Kolodny provides a

called “cognitivists” about practical rationality hold that instances of practical inconsistency or incoherence necessarily involve doxastic inconsistency or incoherence. These proposals have received a great deal of attention, but I am not persuaded that they are particularly plausible. Even if we were to grant that practical rational failures are undergirded by epistemic rational failures (which, I think, is unlikely), the cognitivist would have to provide an explanation of the normativity of epistemic rationality.

second potential route to the conclusion that inconsistency in intention involves failing to respond to the demands of our reasons. This alternative explanation is much more complicated than the first, and the costs of exploring it far outweigh the benefits. I do not want to dwell here on whether Kolodny can provide us with an error-theoretic derivation of every structural requirement. Kolodny has dedicated a number of articles to this endeavor, and while I have some doubts about the success of the project, I think that it would be a mistake to focus my attention on this part of the theory. Plugging through a series of counterexamples to Kolodny's proposed derivations would not demonstrate any deep, principled problem with the theory.

In this section, I have aimed only to explain how Kolodny's new error theory is intended to work. For the purposes of this chapter, I am content to suppose that the new error theory could locate a narrow-scope, reasons-relative failure in any agent who possesses a combination of attitudes that we would customarily associate with structural irrationality. I believe that even if the error theory were to succeed in this respect, it would still face serious problems. In Sections 3.6 and 3.7, I will focus on a pair of problems that together cast considerable doubt on the project's fundamentals. In order to pave the way for these objections, I must do some set-up work in Section 3.4.

3.4 – The Mary-Mary* Case

As I have just explained, the new error theory proposes that the normative failure in a case of apparent structural irrationality is in fact a failure to do what is required by some non-structural requirement of reason (such as the requirement to not hold beliefs for which one has insufficient reason). But there are counterexamples that cast doubt on the plausibility of this explanatory strategy. Specifically, there are cases in which an agent's combination of

attitudes seems to make her do better than her peers in a certain respect, even though she is blatantly doing a worse job of responding to the requirements of her reasons. Cases like these suggest that a wholly reasons-based account is missing something.

Consider two agents, Mary and Mary*, who are found in nearly identical physical and psychological circumstances, in qualitatively identical though separate universes. Mary and Mary* are each aboard a sinking ship in the middle of the ocean, scrambling to find a way to save themselves and the other people aboard. Mary and Mary* are aware of two potential courses of action, each of which seems extremely important: first, a flare can be shot into the sky which would alert passing vessels to their distress; second, an emergency raft can be lowered into the ocean so that passengers can begin to leave the sinking ship. However, Mary and Mary* each firmly believe that it is impossible to both fire a flare into the sky and release the emergency raft into the water. As far as each believes, their ship will be completely submerged in less than a minute, and there is simply no time to both fire a flare into the sky and release the emergency raft. Mary and Mary* believe that the flares are stored on the west-facing end of the ship, that the emergency raft is located on the east-facing end of the ship, and that accomplishing either task will take the entire amount of time remaining before their ship has gone underwater. Mary and Mary* see themselves presented with two extremely important though mutually unrealizable ends.

In fact, Mary and Mary* are mistaken: it is possible to fire a flare into the sky and release the emergency raft into the water. Immediately next to the crank that lowers the emergency raft on the east end of the ship, there is a button that automatically discharges the flare cannon located on the west end of the ship. Mary and Mary* were told about this button multiple times during their safety lessons, and they even recently passed tests demonstrating that they had internalized the information about the flare-button's placement.

Even more, there are signs strategically positioned around the ship that expressly state the button's placement next to the escape raft's crank. But Mary and Mary*, due to some combination of stress, forgetfulness, and inattentiveness, falsely and unjustifiably believe that the two compatible ends are mutually unrealizable. Not only are these ends compatible, but they are both ends which the Marys have decisive reason to pursue. In this situation, Mary and Mary* each have decisive reason to intend to fire a flare into the sky and decisive reason to intend to lower the emergency raft into the water, although they each believe that doing both is impossible.

As described so far, Mary and Mary* are in identical circumstances. But now I will amend the story to add one crucial difference between the two agents. Suppose that Mary, in believing that the two ends are incompatible, holds only one relevant intention: she intends to lower the raft into the water. (She figures that it is better to be stranded on an emergency raft without having alerted anyone than to successfully alert passing ships but drown before they arrive.) But Mary*'s attitudinal situation is different. Mary* both intends to fire a flare into the sky and intends to lower the raft into the water, despite her belief in the impossibility of doing both things together. In short, Mary* possesses a combination of attitudes that we would normally associate with structural irrationality, while Mary does not.

In light of Kolodny's error theory of structural rationality, what can we say about the normative circumstances of Mary and Mary*? First, consider Mary*. Mary* possesses a combination of attitudes which we would normally associate with structural irrationality: she has two intentions which she believes are incompatible. According to Kolodny's error theory, Mary*'s mistake does not reside in her harboring a rationally impermissible combination of attitudes, for there are no such impermissible combinations. If the error theory is correct, then it must be that Mary* is not doing right by all of her reasons. And,

indeed, we are able to find the kind of mistake that Kolodny's theory anticipates: Mary* has a belief for which she has insufficient reason. As explained above, Mary* has a false and unjustified belief in the incompatibility of firing the flare and lowering the raft. Otherwise, Mary* is doing quite well by her reasons; after all, she has the two intentions which she has decisive reason to hold.

This may sound fine so far, but the error-theoretic explanation of Mary*'s situation looks a bit strange when viewed alongside an analysis of Mary's. Like Mary*, Mary has a false and unjustified belief in the incompatibility of the two relevant actions. Although Mary is not exhibiting the kind of attitudinal arrangement that we would normally associate with structural irrationality, she is making the same mistake which the error theory finds in the structurally irrational agent. Moreover, Mary is making an additional normative mistake: she is failing to hold an intention (to fire a flare into the sky) for which she has decisive reason. It seems, then, that Mary is doing doubly bad by her reasons: first, she has a belief for which she lacks sufficient reason; second, she lacks an intention for which she has decisive reason. Mary is normatively worse-off than Mary*, as Mary* is only failing in one respect.

This is a bizarre result. If we only look at their reasons—which the new error theory would have us do—then it is correct to say that Mary is doing worse than Mary*. But what this case reveals is that there is something important about these agents' normative circumstances which is not captured by a wholly reasons-focused analysis. It seems undeniable that Mary* is making a mistake that Mary is not making. Specifically, Mary* seems to be making the mistake of *holding intentions despite her belief in their inconsistency*. Although Mary is flouting reasons that Mary* is not, at least Mary is not making this *other* mistake. We might say “At least Mary is not being *irrational*.” This indicates that there is something independently deficient about the structural arrangement of Mary*'s attitudes. It

is not enough that the new error theory can find *a* problem in Mary*'s case: it must be able to find the *right* problem. And Mary's* problem, intuitively, is not simply a failure of responding to reasons.

3.5 – Enkrasia and the New Error Theory

The error theory does not lack the resources to provide a more complete analysis of the Mary-Mary* case. Kolodny is aware of the kind of issue I am raising, and he believes that the new error theory can accommodate the phenomenon I am gesturing toward.⁴⁶ The proposed solution is to appeal to the intuitively plausible idea, outlined in Section 3.2, that rationality demands *enkrasia*—that we are required to respond appropriately to our perceived reasons. While Raz and Kolodny originally thought to build an account of rationality entirely on the basis of enkratic principles, Kolodny's newer strategy is to use this idea only as a supplement. This is a more plausible place for enkratic requirements, as it requires them to do considerably less work in the error theory. Instead of expecting enkrasia to hold the whole weight of the error theory, it is only needed to do some mop-up work.

According to the augmented version of Kolodny's new error theory, someone like Mary does well (in a way that Mary* does not) by responding appropriately to her perceived reasons. We can imagine that Mary, in believing (falsely) that the two intentions could not be mutually realized, thereby judges that she has insufficient reason to hold the two intentions together. Mary might think: "I'll never have a time to do both, and so it would be absolutely dangerous and wrong of me to try!" Suppose that this is why Mary holds only one of the two intentions. Although Mary's judgment about her reasons is incorrect, she is responding

⁴⁶ Kolodny 2008a, pp. 32-33.

appropriately to the reasons she believes apply to her. Assuming that Mary* has the same judgment about her reasons, she is being insensitive to those perceived reasons when she nevertheless intends to fire the flare and intends to lower the raft. This is the spot where Mary* is failing and Mary is succeeding.

This addition to Kolodny's new error theory leaves the core of the theory in place, but adds a second device: while some of our intuitions about structural rationality are explained away by the insight about failing to respond to actual (narrow-scope) requirements of reasons, other intuitions are explained away by appealing to the demands of enkrasia. In light of this amendment, we can now summarize the normative circumstances of Mary and Mary* as follows: Mary and Mary* each have a belief for which they lack sufficient reason (i.e., in the incompatibility of the two relevant actions); Mary, but not Mary*, fails to have an intention for which she has decisive reason (i.e., the intention to fire a flare); Mary*, but not Mary, fails to respond appropriately to her perceived reasons (i.e., by having the pair of intentions despite her judgment). Mary and Mary* each get something right that the other does not. The amendment to the new error theory promises not only to find a problem in Mary*, but also the right kind of problem.

The Mary-Mary* case is a helpful device for illustrating the two-part mechanics of the new error theory. And looking at this case places us in a good position to begin revealing the account's inadequacies. Before moving on to examine the most pressing challenge facing the error theory in Section 3.6, I would like to briefly highlight one problem with Kolodny's new account that results from the diminished but important role it gives to enkrasia.

Recall that the content problem of structural rationality demands that we explain what makes a rational requirement count as a rational requirement (rather than a member of some other domain) or, to put the point another way, what unifies rationality as a set of

requirements. Of course, the error theorist will reply that rationality *has* no content and that the set of normative rational requirements is *empty*. There is, if we take the error theorist at her word, nothing to unify. But the error theorist nevertheless owes us an explanation of why it *seems* to us that rationality is a unified subject matter.

On its face, the original error theory of Raz 2005a and Kolodny 2005 offers an attractive and persuasive account of rationality's apparent unity. According to the old error theory, every requirement of rationality is grounded in the fundamental demand that we respond appropriately to our perceived reasons. If the original error theory's derivation scheme were successful, then this would explain why rationality appears to be a unified subject matter: disparate cases of apparent structural rationality appear unified to us because every irrational agent is, in fact, failing to respond appropriately to her perceived reasons. Unfortunately, it turns out that it is implausible that every violation of a rational requirement involves a failing of *enkrasia*. So rationality cannot be unified in this neat and tidy way.

The two-part mechanics of Kolodny's new error theory are superior in that they promise to locate a problem in every problem-case: every agent who seems to be suffering from irrationality has her irrationality accounted for within the two-part system. But this extensional success comes at the cost of destroying the unity of the original error theory. If the new error theory were correct, then in some cases, an agent with such-and-such an arrangement of attitudes will seem irrational to us in virtue of her response-failure to actual reasons; in other cases, an agent with such-and-such an arrangement of attitudes will seem irrational to us in virtue of her response-failure to her perceived reasons. But it is not clear to me that this kind of explanation is faithful to our intuitions about rationality. Our intuitions about rationality do not have this bipartite structure. The requirements of *enkrasia* seem closely connected to other requirements of rationality, just as requirements concerning

consistency in belief seem connected to requirements concerning means-end coherence. The new error theory cannot easily accommodate the fact that our judgments about rationality suggest that rationality is unified.

This observation is suggestive rather than decisive, and so I do not want to make too much hang on this point. If the new error theory were otherwise successful, perhaps its fragmented mechanics would not be a strong reason to discount it.

3.6 – The Normativity of Enkratic Requirements and the Normativity of Rationality

I now turn to the most pressing problem facing the error theories, both new and old: the normative status of enkratic requirements. As we have seen, both the old Razian error theory and Kolodny's new error theory rely crucially on requirements demanding that we respond appropriately to our perceived reasons. Enkratic requirements seem to have all of the markings of standard requirements of structural rationality: they demand that we not have particular combinations of attitudes. While an error theory is free to appeal to these requirements, doing so creates a new explanatory burden. An error theory cannot explain away some structural requirements by appealing to other structural requirements, unless it is prepared to also explain the status of these other structural requirements.

Of course, Raz and Kolodny want deny that the requirements of enkrasia are legitimate normative requirements. They will hope to persuade us that the requirements of enkrasia are themselves but apparent requirements, which can be explained away like all other requirements of structural rationality. The question is whether either account has the resources to discharge any commitment to legitimate requirements of enkrasia.

It is worth making the obvious point that new error theory's primary method of explaining away structural requirements is ineffective in the case of enkratic requirements. It

is false that whenever one goes wrong by one's perceived reasons one is thereby failing to conform to some of one's actual reasons. Consider an agent who falsely—but on the basis of compelling evidence—believes that she has decisive reason to intend to X, while at the same time she does not intend to X. This agent need not be lacking any attitude for which she has decisive reason, as she may have no reason at all to X (or intend to X). And it may be that the agent does not hold any attitude for which she has insufficient reason, as the available evidence may strongly (though misleadingly) support her belief that she should intend to X. So the requirements of enkrasia cannot be subjected to the new error-theoretic explanation that Kolodny favors. This means that the old error theory and Kolodny's new error theory are in the same spot with respect to the requirements of enkrasia.

The going error-theoretic strategy is to suggest that the explanatory burden is much lower in the case of enkratic requirements than for other requirements of rationality. In the case of some structural requirements, it has proven to be quite complicated to explain why we believe that we are bound by these requirements even though they do not exist. But the error theorist argues that enkratic requirements are different—that there is no puzzle about why it seems to us that we are bound to comply with the demands of our reasons. “Why does it seem to you that you are required to do that which it seems to you that you are required to do?” is barely intelligible as a question. Of course it will seem to you that you are required to comply with your reasons! This is thought behind the so-called *transparency account* first proposed in Kolodny 2005.⁴⁷

We can return to the Mary-Mary* case to illustrate the idea. When Mary* fails to

⁴⁷ See, also, Kolodny 2008a and Kolodny 2008b. In both articles, Kolodny reinforces that enkratic requirements must be understood as mere *seeming* requirements, rather than actual, normative requirements.

comply with her judgments about her reasons—by simultaneously intending to fire the flare and lower the raft, despite her belief that her reasons require her not to pursue both courses of action at the same time—it will seem to her, first-personally, that she is doing something wrong. By Mary*'s own lights, her simultaneous pursuit of both ends will seem like a dangerous mistake. And, indeed, this is all that a failure of enkrasia amounts to; a failure with respect to one's normative judgment just *is* a failure by one's own lights. So when we assess Mary's and Mary*'s normative circumstances, and we judge that Mary* is making a mistake that Mary is not making, the error theorist will suggest that we are only noticing that Mary* is making a mistake by her own lights. If this is so then (so goes the theory), we do not need to appeal to a full-blown, normative requirement of rationality governing enkrasia. On this way of thinking, the requirements of enkrasia are themselves but apparent requirements, the apparent force of which is provided by our apparent reasons.

This is what the error theorist would have us believe, but the story elides an important feature of the phenomenon of enkratic failure. It is true, of course, that Mary* is making a mistake by her own lights. And, first-personally, in the heat of the moment, it may be this perception of substantive error (e.g., “I know I shouldn’t be trying to do both of these things!”) which dominates Mary*'s mental experience. But this observation should not distract us from the fact that Mary* is, in addition to making a mistake by her own lights, making an *actual mistake*. Mary*'s actual mistake is the mistake of making-a-mistake-by-her-own-lights. This may sound silly, but it is an important point: a failure to do right by one's own lights is a failure of the not-just-by-one's-own-lights variety. The driving intuition behind the Mary-Mary* case is precisely that it is an objective rational failing to not do what one believes one's reasons demand.

The error theorist directs our attention to the idea that, first-personally, it is easy to

explain why we feel bound to comply with our perceived reasons. This slight-of-hand maneuver masks the fact that, third-personally, we routinely assess other agents as bound to comply with the reasons they take themselves to have.⁴⁸ But it also ignores the complexities of our self-evaluations. Suppose that I believe that I have overwhelmingly strong reason to stop eating boxes of donuts before bed, but that I nevertheless persist in eating them each night. As I set out toward the kitchen for tonight's treat, I might think "I am such a failure for *eating another donut*," where my focus is on the substantive mistake of eating another donut. At the same time, however, I might also think "I am such a failure for eating another donut *even while I know I shouldn't*," where my focus is on the structural mistake of failing to comply with my own normative judgments. If an omniscient dietician appears and informs me that (contrary to all our worldly evidence), eating donuts before bed is the secret to human health, she will ease my guilt about only one of my two failures. Even from the first-person perspective, the demands of my reasons are separable from the demands of enkrasia.

The Mary-Mary* case initially seemed problematic for the error theorist because it seemed to reveal that although the error theory could account for a problem in Mary*'s case, it could not account for the right problem. The error theory's appeal to requirements of enkrasia was intended to patch up this hole in the account. This left the error theory saddled with a new set of structural requirements which demand explanation. An error-theoretic account of enkratic requirements may be possible, but the strategy just explored is not a

⁴⁸ Kolodny's transparency account provides an attempted explanation of our third-personal judgments about others who violate enkratic requirements. According to this view, if I see an agent failing to intend to X despite her belief that she has decisive reason to intend to X, my critical appraisal would be limited to the judgment that she should, by her own lights, intend to X, and that she will be making a mistake, by her own lights, unless she does so. My contention is that this strategy only captures a piece of our common-sense thinking about enkratic failures.

winning one. We cannot appeal to enkratic requirements to do work which the error theory could not otherwise do and then declare that they are nothing but illusions provided by the nature of first-person deliberation.

The error theorist will be forced to cede some ground here. He may admit that there are requirements of enkrasia, but while denying their normative significance. To make this move the error theorist could appeal to a distinction first noted by John Broome (albeit, for very different purposes). Broome draws our attention to the fact that we use the word “requirement” in at least two different senses. He distinguishes “property requirements” from “source requirements.”⁴⁹ A source requirement is a requirement which makes an active demand on one’s conduct or behavior, and whose authority derives from some source. Moral requirements and legal requirements are, arguably, examples of source requirements. And these requirements are legitimate, robustly normative requirements of the sort that I have been analyzing throughout this project. Property requirements, on the other hand, are merely conditions which an individual must satisfy in order to have a specified property. The requirements of smugness or beauty are, arguably, mere property requirements. There is no denying that there are requirements of smugness and beauty: we often make statements like “smugness requires a feeling of self-satisfaction” and “beauty requires gracefulness.” But in saying such things we do not commit ourselves to the idea that certain sources—Smugness and Beauty—place demands on our conduct. We do not think that the requirements of smugness or beauty call on us to be or act a certain way. In other words, the requirements of

⁴⁹ See Broome 2013, Chapter 5.

smugness and beauty are not normative requirements.⁵⁰

The error theorist could attempt to use this distinction to insulate his appeal to enkratic requirements. The error theorist may conceive of rationality (or this part of rationality) in purely definitional or descriptive terms. It may be that an agent who violates the (property) requirements of enkrasia is simply failing to instantiate the property of rationality, much the same way that the ugly person fails to instantiate the property of beauty. And this could simply mean that she is not functioning in the way that agents typically function *qua* rational agent (i.e., by responding to their own normative judgments). The error theorist may admit that it is a central part of being a rational agent that one responds to one's judgments about one's reasons, without thereby accepting that there are normative requirements demanding that we do so.

I suspect that some version of this thought is operating in the background of Raz's seminal articles on instrumental rationality. While Raz freely writes of agents' rationality and irrationality, and of the poor functioning exhibited by irrational agents, he avoids saying that there are any *requirements* of rationality. Raz's choice to treat rationality solely in property terms is rhetorically wise, as it avoids raising questions about the source and normativity of these requirements. Raz is happy for us to see rationality (i.e., responsiveness to perceived

⁵⁰ In many cases, property requirements correspond to source requirements in that you can only have a specified property if you satisfy a certain range of source requirements. For example, you can only have the property of being moral if you satisfy the (source) requirements of morality. While every set of source requirements corresponds to a potential property (which one might have in virtue of satisfying the source requirements), not every property requirement corresponds to a set of source requirements. For example, although there are property requirements of smugness (i.e., conditions that must be met in order to have the property of being smug) there are no source requirements of smugness.

reasons) only as a property of agents, rather than as a source of binding requirements.

Unlike Raz, Kolodny does appeal to explicitly labeled requirements governing enkrasia. But I suspect that he would be satisfied to conclude that these requirements are mere property requirements. This move would produce a somewhat messy though internally consistent version of the new error theory. For many apparent structural requirements (e.g., the requirements of belief consistency, intention consistency, means-end coherence, and so forth), the new error theory acknowledges that there is an underlying normative phenomenon at work, but explains the requirements away by a locating non-structural, reasons-based failing in relevantly situated agents. Then, in order to explain away our judgments about the distinctive rational failings of people like Mary*, the new error theory allows that there are independent requirements of enkrasia. But the theory asserts that enkratic requirements are only “requirements” in a very loose sense of the word. They are, according to the proposal being canvassed, mere property requirements—conditions that agents must satisfy in order to fully instantiate a property we call “rationality.”

This appeal to the distinction between property and source requirements may be the most plausible strategy available to the error theorists, given their reliance on enkrasia. But I find the response deeply unsatisfying. I am sympathetic to the idea that an enkratic failure involves a failure to function normally or appropriately as a rational being. Indeed, it is quite plausible that every failing of structural rationality involves a related form of functional failure. As I will go on to explain in the coming chapters, I suspect of our intuitions about rationality and irrationality are undergirded by judgments about the standards of proper agential functioning. So I am not hostile to Raz’s thought that irrationality involves a functional deficiency. I am hostile only to the idea that we might move directly from this observation to the conclusion that rationality is a mere property, rather than a source of

normative requirements.

We do not conceive of the irrational person as merely failing to instantiate a property, as we would the non-beautiful person or the non-smug person. The driving intuition behind the Mary-Mary* case is that Mary* is getting something interestingly *wrong*. If irrationality is a form of faulty functioning, as Raz would have us believe, then it is a form of faulty functioning which we *ought* not to exhibit. Of course, as we have seen, there are serious problems with interpreting the “ought” I am employing here as the familiar, reasons-based “ought.” But I am reminding us that, intuitively, rationality is rather clearly a source of requirements, rather than a mere property. To do its job well, the error theory must be able to explain, or explain away, these central aspects of the phenomenon of irrationality. Categorizing certain rational requirements as mere property requirements is a bold and revisionary move, not an accommodation of our starting-point intuitions.

If we arrive at a range of apparent normative requirements which the error theory cannot explain away, it is philosophically illicit to declare by fiat that those difficult requirements are simply not normative. After all, we could have jumped directly to this stage as soon as we noticed in Chapter 1 that we do not have any general reasons to comply with the requirements of structural rationality. We could have declared at the outset that all of the apparent requirements of structural rationality—the requirements of belief consistency, intention consistency, means-end coherence, enkrasia, and so forth—are but conditions that an agent must satisfy in order to count as a normally functioning rational being, rather than normative demands. Just as it would have been inappropriate to make this move at the beginning, it is inappropriate to make this move here. We should not so quickly conclude that the requirements of enkrasia, or the requirements of rationality more generally, are mere property requirements.

Of course, the error theorist has the option of offering a new error-theoretic explanation of enkrasia, or a different error theory of structural rationality altogether. But I know of no alternative approach for the reasons fundamentalist to take. So unless a new explanation is offered, I think we must conclude that the error theories have failed.

3.7 – The Failure of the Error Theories

I have used this chapter to see how far we can take the reasons-based error theories, but I must conclude that they are unsuccessful. Both Raz's original error theory and Kolodny's new error theory face an uncomfortable choice. They could simply refuse to provide us with a sufficient explanation of our intuitions about irrationality and enkratic failure. But this would be tantamount to admitting the failure of the theory. Alternatively, they could entertain the possibility that the principles of enkrasia are legitimate, non-reasons-based normative requirements. But this would involve abandoning the entire error-theoretic project.

Unsurprisingly, abandoning the error theory is precisely what I think the reasons fundamentalist should do. As we have seen, the error theory culminates in a bald refusal to consider the possibility of non-reasons-based normative requirements. In the following chapters, I will suggest that we have good grounds for seeing the requirements of structural rationality as legitimate, normative requirements, even if they are not requirements with which we have reason to comply. My arguments will be aided by Raz's suggestion that irrationality is a matter of improper functioning.

4 – THE POSSIBILITY OF NORMATIVE PLURALISM

4.1 – Toward a New Account of Structural Rationality

I will begin the second half of this project by recounting where we have been and outlining where we are headed.

In Chapter 1, upon reflecting on our common-sense judgments about agents with seemingly improper sets of intentions and beliefs, I introduced the requirements of structural rationality: wide-scope prohibitions on some combinations of attitudes. At first blush, these requirements seemed almost undeniable, as they made sense of our judgments about rationality and irrationality. But difficulties arose when we learned that it is implausible that we have any sufficiently general and persistent reasons to comply with these apparent requirements. This finding called into question the normativity of rationality, as well as the unity and content of the rational domain. We were left wondering whether there could be any way to explain and justify the requirements that initially seemed so intuitively plausible.

In Chapter 2, I took a close look at Christine Korsgaard's constitutivism. Korsgaard offers a theory of rationality, though one that disagrees with my foundational claim that rational requirements are wide-scope prohibitive principles. While I argued that Korsgaard's view, as stated, could not provide an account of the requirements of structural rationality, I harvested two important results from the discussion. First, I found that, regardless of whether there may be some requirements on processes of reasoning or thinking, we still need an account of purely structural requirements. Second, and more positively, I found that Korsgaard's arguments suggest how a requirement's normativity might not be based in any facts about reasons.

In Chapter 3, I looked at the recent error theories of the reasons fundamentalists. These error theories attempt to explain away the requirements of structural rationality by appealing (to various degrees) to the demands of *enkrasia*. I found these views to be unsuccessful, because they are left with an uncomfortable choice between either losing their error-theoretic status or leaving certain core intuitions about rationality unaccounted for. At the same time, I registered my agreement with the idea, first explored by Raz, that irrationality involves a form of improper functioning as an agent.

The rest of this project—this chapter, along with the two following—is dedicated to offering an account of structural rationality that explains both its content and its normativity. In providing this account, I will rely in various ways on the views that I have just considered and rejected. I will offer a (selective) synthesis of ideas from both the constitutivist and the reasons fundamentalist, with the goal of showing that there are, as intuition suggests, wholly normative requirements of rationality which take the form of wide-scope, state-based prohibitions on our combinations of attitudes. I will argue that while there are requirements whose normativity is constituted by reasons (as the reasons fundamentalist would suggest), there can also be non-reasons-based normative requirements, as the constitutivist has indicated. I will argue that the requirements of structural rationality are examples of non-reasons-based normative requirements, and I will claim that these requirements are rooted in rational agents' agential powers.

In order to build this positive account, however, I must first attend to a central claim on which my proposal depends and which places me in deep disagreement with both the reasons fundamentalist and the constitutivist. The views I have considered so far in this dissertation are *monist* theses about normativity. According to the reasons fundamentalist, all normative claims ultimately bottom out in claims about reasons. And according to the

constitutivist, all normative claims ultimately bottom out in claims about agency. Although these views disagree with each other about what gives rise to and binds together the members of the normative domain as a whole, they agree that there is some basic normative commodity from which all other normative goods are derived. According to both of these views, there is only one ultimate source of normative requirements.

Against these monist views, I will defend a thesis that I call *normative pluralism*. The thesis, stated concisely, says that there is more than one fundamentally distinct type of normative requirement or, put slightly differently, that there are legitimate requirements made relative to fundamentally distinct normative standards. Normative pluralism is not itself a view about the requirements of structural rationality, but instead a view about the foundations of normativity in general. One could be a normative pluralist while rejecting everything that I will go on to say about structural rationality. But the account of rationality that I favor relies on the truth of the pluralist thesis, and so I must defend the pluralist proposal before offering my account of structural rationality. Defending normative pluralism is the focus of the present chapter.

I will defend normative pluralism by arguing that a requirement is normative for an agent if and only if the agent's violation of the requirement would be a *criticizable failure*. Pluralism's truth is secured by the fact that there are fundamentally distinct and independent grounds on which a lack of compliance with some putative requirement can constitute a failure. For instance, although some principles bind us because if we do not comply with them we will be failing to do what our reasons demand, other principles are binding because if we do not comply with them then we will be exhibiting some independent agential defect. There can be, I will argue, principles which it is criticizable to violate even if we have no independent reason to comply with their demands.

After I have offered my defense of normative pluralism in this chapter, I will return to the topic of structural rationality. In Chapter 5, I will begin to make the case that the requirements of structural rationality are normative because failing to comply with them is a criticizable agential failure. But the progress of my argument in the fifth chapter will be somewhat indirect. Rather than straightforwardly arguing that structural irrationality is criticizable on non-reasons-involving grounds, I will set out to provide a comprehensive characterization of the *content* of rationality. I will argue that the behavior we identify as structural irrationality always involves the same form of agential self-undermining. Although I cannot yet clarify exactly what this means, I will argue that every apparent requirement of structural rationality—IC, BC, EI, ME, and so on—is but an instance of a more general demand that we not undermine the success of our own beliefs and intentions. I will develop a principle that I will call *the General Requirement of Structural Rationality*, and I will propose that this General Requirement allows us to explain the unity of the rational domain and to articulate its members. In the sixth and final chapter, I will complete my account of rationality by reflecting on why the relevant form of attitudinal self-undermining is a criticizable agential failure, and I will highlight a number of questions that, though beyond the scope of this project, deserve further attention.

But in the current chapter I will pursue only the first of these goals: I will seek to sketch a preliminary defense of normative pluralism. In the process, I will occasionally take note of the requirements of structural rationality. But my focus in this chapter is on pluralism itself.

4.2 – Defining “Normative Pluralism”

I must begin by clarifying what I mean by “normative pluralism” before I can tell you why

you should find the view plausible.

Normative discussions are discussions about what we are required to do, or, in other words, about what we ought to do. The words doing the normative work here are ‘requirement’ and ‘ought.’ If there were no facts about what we ought to do, then there would be no normative facts at all. So I will assume that the most familiar and uncontroversially normative term is the term ‘ought.’ And, similarly, I will take it that the most familiar and uncontroversial type of normative claim is what I will call an ‘ought’-claim: a claim like *You ought to X*.⁵¹ So I will characterize normative pluralism by looking at what it has to say about ‘ought’ and about ‘ought’-claims.⁵²

A quick, though potentially misleading, way of characterizing normative pluralism is to say that, according to the pluralist view, there is more than one legitimate sense of ‘ought’ and more than one type of true and binding ‘ought’-claim. This is the characterization of normative pluralism that I offered in the introductory section of this chapter. But this way of putting things may be misleading because it may make normative pluralism sound like a position that nearly everyone already accepts. After all, it is not uncommon for normal people—or philosophers, anyway—to speak as if there are different kinds of ‘ought’-s.

For example, one might say that there are legal ‘ought’-claims and moral ‘ought’-

⁵¹ Not all claims of the form ‘You ought to X’ are legitimately normative. Many ‘ought’-claims which purport to be normative are not normative, because they are false. Differently, some claims of the form ‘You ought to X’ do not purport to be normative at all. For example, one might say “According to the news, it ought to rain tomorrow,” where the ‘ought’ is used in a predictive rather than a normative sense. I am focused only on normative (and purportedly normative) ‘ought’-claims.

⁵² I do not mean to suggest that ‘ought’ is the only foundational normative term or concept. I believe that there are other equally fundamental normative terms, such as the term ‘requirement.’

claims and prudential ‘ought’-claims, and so on. I ought to file my taxes by the fifteenth, and to say so invokes a legal sense of ‘ought’ and makes a legal ‘ought’-claim. I ought to keep my promise to water your plants today, and to say so invokes a moral use of ‘ought’ and makes a moral ‘ought’-claim. I ought to exercise tonight, and to say so is to make use of a prudential ‘ought’ and to make a prudential ‘ought’-claim. This is not an altogether unfamiliar way of speaking. So it may look as though we already all agree that there is more than one sense of ‘ought’ and more than one legitimate type of ‘ought’-claim. Despite appearances, this way of speaking does not involve an endorsement of normative pluralism. Once we understand what I mean by “type,” it will be clear these three ‘ought’-claims are most plausibly all claims of the *same* type. So I need to say more about what I mean when I say that normative pluralism is the view that there are ‘ought’-claims of different types.

In the relevant sense of “type,” two ‘ought’-claims are of the same type when they are claims that could, in principle, *conflict* with each other. Two ‘ought’-claims conflict with each other in some circumstance if the truth of one claim (together, perhaps, with certain other facts) would ensure the falsity of the other claim. In other words, two ‘ought’-claims can conflict with each other if they may be in competition with each other for their truth. In the above examples—of paying taxes and fulfilling promises and exercising for one’s health—all of the involved ‘ought’-claims are most plausibly claims of the same type, because these ‘ought’-claims could conflict with each other.

Imagine that it is late in the afternoon on the fifteenth of the month and I have only enough time to file my taxes, water your plants, or exercise. For the sake of argument, just allow that I cannot, no matter how hard I were to try, do any more than one of these three things tonight. In such a case, it would be unusual to hold that each of the three ‘ought’-claims would remain true. It sounds unnatural to say that I ought to file my taxes and that I

ought to water your plants, even though I no longer can do both. This is because the ‘ought’-claims conflict with each other, as either claim’s truth would guarantee each other claim’s falsity. Perhaps what I ought to do in these circumstances is water your plants. If this is true, then it is guaranteed to be false that (tonight) I ought to exercise and false that (tonight) I ought to file my taxes. So I believe that claims like ‘I ought to file my taxes,’ ‘I ought to water your plants,’ and ‘I ought to exercise’ are all ‘ought’-claims of the same type and all invoke the same sense of ‘ought.’ If the thesis of normative pluralism is true, then there are some ‘ought’-claims which cannot conflict in this way, because the ‘ought’-claims are of fundamentally different types.

I must explain, then, why some ‘ought’-claims can conflict with each other while others cannot. The answer is that different types of ‘ought’-claims are made relative to different *standards* or, in other words, are made regarding fundamentally different normative subject matters. Consider, again, the examples of fulfilling a promise, doing taxes, and getting some exercise. Each of the involved ‘ought’-claims can conflict with each other because they are each made relative to the same ultimate standard and involve the same fundamental normative subject matter. Specifically, in my view, each of the three ‘ought’-claims is made relative to the standard of most or decisive reason.

As the reasons fundamentalist would expect, these ‘ought’-claims are each made true (when they are true) by the reasons I have for pursuing the various courses of action. If in some circumstance I ought to water your plants, ought to file my taxes, and ought to get some exercise, then this is because I have decisive reason to do each of these three things. The ‘ought’-claims can conflict with each other because I cannot have decisive reason to take more than one of various incompatible courses of action. When time prevents me from completing more than one of the actions in question, there is a single, all-in fact about which

of these three things I ought to do, made relative to the reasons standard. If this is correct, then although we might sometimes speak of moral, legal and prudential ‘ought’-claims, we do so without thereby endorsing any form of normative pluralism.⁵³ To be normative pluralists, we would have to believe that there are different types of ‘ought’-claims made relative to fundamentally distinct normative standards.

Although I suspect that no one would hold such an implausible view, it is worth flagging that one *could* believe that moral, legal, and prudential ‘ought’-claims are each made relative to fundamentally distinct standards and, as such, that they are different types of ‘ought’-claims. Holding such a view would make one a kind of normative pluralist. This view would imply that morality, the law, and prudence are all normatively isolated subject matters, and that there are no all-things-considered normative facts that take into account these disparate sources of normative claims. In the case above, where it is late on the fifteenth of the month and I only have time to pursue a single course of action, this view would allow that I nevertheless ought to water your plants, ought to file my taxes, and ought to exercise. The view would insist that these three claims employ different forms of ‘ought’ and that, as such, there is no all-things-considered fact about which of these three things I “really” ought to do. This kind of view is not incoherent, although it conflicts with common sense. I have

⁵³ According to the line of thought I am employing in this paragraph, to say that I “ought legally” to X is only to say that a certain subset of reasons having to do with the law, if held in isolation from reasons that do not have to do with the law, decisively count in favor of my X-ing. According to this position, all reasons are ultimately subjected to the same overall normative ‘weighing’. When the weighing is done, all reasons are *just reasons*, and there is a single all-in standard of most or decisive reason which takes into account legal reasons, moral reasons, prudential reasons, aesthetic reasons, familial reasons, and so on. On this view, there is in fact only a single type of ‘ought’-claim at play here.

been assuming that this view is false in order to help bring out what is distinctive about normative pluralism. But it is not important for my purposes that this weird view is false, because the truth of the view would itself establish the truth of normative pluralism.

Normative pluralism is the thesis that there can be more than one type of legitimate and binding ‘ought’-claim because ‘ought’-claims can be made relative to distinct normative standards. While all pluralists must accept this claim, there is no need for them to agree about anything beyond this. There are innumerable conceptually available varieties of normative pluralism. In the remainder of this project, I will discuss only a very small range of potential versions of pluralism. As I have just made clear in my use of examples, the version of normative pluralism that I favor takes onboard the reasons fundamentalist’s claim that there are legitimate, non-derivative ‘ought’-claims made relative to the reasons standard. But I am a normative pluralist because I believe that there are other legitimate types of ‘ought’-claims which are made relative to a fundamentally distinct standard. In principle, one could be a normative pluralist who rejects the reasons fundamentalist position entirely. The boundaries of normative pluralism are broader than my own theoretical allegiances.

Although the bare pluralist thesis is not itself a view about rationality (or any other substantive matter), it is obvious how one might offer an account of the normativity of structural rationality that relies on a version of normative pluralism. One could argue that philosophers have made the mistake of assuming that the ‘ought’ operative in the requirements of structural rationality could only be the ‘ought’ of reasons, when in fact there is a distinct, non-reasons-involving standard according to which the ‘ought’-claims of structural rationality are made. This is precisely what I will argue in Chapters 5 and 6.

I have a rough sketch of the pluralist thesis on the table. I have not yet offered any particular pluralist proposal. And so the obvious question is: what other legitimate uses of

‘ought’ are there and what could be their basis? But before offering a defense of the version of normative pluralism that I find most plausible, I will say something about the status of the view in contemporary normative philosophy.

4.3 – The Current Status of Normative Pluralism in Philosophy

As I suggested in Section 4.1, the views I have discussed in the first three chapters of this project are all monist, rather than pluralist, views about normativity. According to the reasons fundamentalist, there is only one ultimate normative standard: the reasons standard. The reasons fundamentalists tells us that all legitimate ‘ought’-claims are made true by the reasons we have for complying with them. The constitutivist agrees that there is only one ultimate normative standard, but holds that this standard is the standard of agency itself. According to the constitutivist, any legitimate ‘ought’-claim is made true because it is appropriately conceptually connected to the nature of thought and action.

Normative pluralism is not endorsed by many (or perhaps any) current writers. I am aware of no single explanation of why this is the case. It may be the result of the ambitiousness of most normative theorists; a comprehensive, single-source account would be an elegant, simple, and powerful way to explain the workings of the normative domain. I suspect that this motivation is shared by reasons fundamentalists and constitutivists alike. At the same time, it may be that monism is simply the received view and that its dominance in contemporary philosophy is more a matter of tradition than argument. Indeed, pluralism is most often not even considered a possibility; it is generally ignored without being argued against.

Although there are very few affirmed normative pluralists within mainstream philosophy, there have been occasions on which the possibility of pluralism has been raised.

In recent years, there have been several cases in which an author's view has appeared to drift toward endorsing or assuming the truth of normative pluralism. In these cases, the author in question makes quick work of issuing a correction. There are two particularly interesting examples of this within the literature concerning rationality.

Recall Joseph Raz's maddeningly difficult article, "The Myth of Instrumental Rationality," which I introduced in Chapter 3. As we saw, Raz argues that even though there is no principle of reasons which demands means-end coherence, there is nevertheless something going wrong with an agent who fails to take the means she views as crucial to the realization of her ends. Raz, making a point which I take seriously and which I will come back to in the next two chapters, suggests that such agents are exhibiting a kind of "faulty functioning." According to Raz, that an agent so situated is defectively functioning is supposed to explain the mistake she is making when she is means-end incoherent. In Chapter 3, I proposed that Raz implicitly advocates for a mere property-requirement conception of rationality, as he attempts to do justice to our intuitions about rationality without endorsing that there are any normative requirements.

Nevertheless, some writers misunderstood Raz and took him to be suggesting that there may be an additional form of normative demand, beyond the normative demands provided by our reasons. Alongside the publication of Raz's paper, David Sobel offered one such interpretation of Raz.⁵⁴ According to Sobel's interpretation, Raz believes that some normative claims have to do with reasons, while others have to do with proper functioning. If this were Raz's actual position, Raz would be a normative pluralist. But Raz's response

⁵⁴ See Sobel 2005.

makes clear his dissatisfaction with Sobel's characterization. He says:⁵⁵

Sobel appreciates [that people who fail to take the means to their ends are exhibiting a functional failure], but he distorts its meaning by thinking that it is a claim about 'two sources of normativity.' [What] I have in mind has nothing to do with different sources of normativity.

This comment lends further credence to the interpretation I offered in Chapter 3. Raz does not advocate that there could be more than one fundamentally distinct kind of 'ought'-claim. To Raz, legitimate 'ought'-claims are, in every case, judgments stating what we have decisive reason to do. Raz's hope is to explain away our intuitions about rationality by appealing to a non-normative standard of proper functioning.

Something similar has happened to John Broome. Broome's case is more confusing than Raz's. Broome used to advocate—or, at least, seemed at one time to advocate—that there can be more than one sense of 'ought' and, accordingly, more than one distinct brand of normative requirement. In his early work on rationality, Broome used a wide-scope 'ought' operator when stating the requirements of rationality. Those requirements appeared extremely similar to the ones I have endorsed in this project, and he called such requirements "normative." Given that Broome at no point argued that we have decisive reason to comply with requirements like these, it was natural to interpret Broome as implicitly suggesting that these principles invoked a distinct form of 'ought.' Some readers, myself included, understood Broome to be attempting to highlight a distinction between the 'ought' of rationality and the 'ought' of reasons. But according to Broome's recent writing, this was not his intention. Broome says that he was insufficiently clear in the past, but that he does not endorse any such pluralist view. According to Broome, to say that one ought to do something is to say the one has decisive reason to do that thing. Recently, Broome has

⁵⁵ Raz 2005b, p. 10.

said that “to say a requirement [...] is normative is to say that the requirement constitutes a reason.”⁵⁶ This is why, as we saw in Chapter 1, Broome is left believing that there must be reasons to comply with rational requirements, even though he can locate no such reasons. Broome, who once seemed to be one of pluralism’s defenders, is in fact an opponent.

It is humorous, and perhaps telling, that both Raz and Broome have been misinterpreted in this way. Of course, given their commitment to reasons fundamentalism, it is no surprise that they would resist being understood as pluralists. But while the reasons fundamentalists are quick to point out that they have not endorsed any pluralist thesis, one could argue that they have no choice. One might argue that reasons fundamentalists, while avowed monists, are in fact normative pluralists in disguise.

Although the fundamentalists argue that all normative claims are constituted by reasons, this simple phrasing ignores the fact that reasons come in at least two very different forms: practical and epistemic. It is not clear how much epistemic and practical reasons have in common with each other beyond their label as reasons. To appeal to the criterion introduced earlier in this chapter, it seems quite likely that an ‘ought’-claim made true in virtue of one’s epistemic reasons and an ‘ought’-claim made true in virtue of one’s practical reasons cannot conflict with each other in the way that normative claims of the same type must be able to. I suspect that this issue is in the background of the ongoing debates concerning whether there are any practical reasons for belief. Imagine a case in which an agent is given overwhelmingly strong practical reason to believe P, even though she has insufficient evidence for P. To the extent that we can make sense of this kind of example, it seems most natural to say that she ought to believe P in the practical-reason-involving sense

⁵⁶ See Chapter 2 of Broome 2013.

of ‘ought,’ but that she ought not to believe P in the epistemic-reason-involving sense of ‘ought.’ It would be odd to expect that the completely disparate epistemic and practical considerations could be combined to generate an all-things-considered normative verdict about what the agent “really” ought to believe.⁵⁷ This suggests that the standard appealed to in the ‘ought’ of epistemic reasons is normatively distinct from the standard appealed to in the ‘ought’ of practical reasons. If this line of thinking is right, then reasons fundamentalists should not have any principled opposition to normative pluralism, as they are already normative pluralists.

Although I find this argument provocative, it does not matter too much for my purposes whether reasons fundamentalists are unintended pluralists. In either case, I am sympathetic to a crucial aspect of the reasons fundamentalist position: I believe that there are basic, irreducibly normative facts about what we have reason to do (or believe, or intend, and so on). And I am content with the idea that many ‘ought’-claims are made legitimate by the reasons we have for complying with their demands. As we have seen so far in this project, however, not every part of the normative world can easily be explained by appealing to our reasons. And if an exclusively reasons-based view prevents us from fully mapping the normative domain, then we must be willing to look beyond the reasons fundamentalist’s system. This is why I am a normative pluralist.⁵⁸

⁵⁷ The reasons fundamentalists may wish to argue that there are no such things as practical reasons for belief. Although this would threaten to make my example irrelevant, it would not demonstrate that the ‘ought’ of epistemic reasons and the ‘ought’ of practical reasons are of the same type. For helpful discussions concerning practical (or pragmatic) reasons for belief, see Williams 1973, Kelly 2002, Reisner 2009, and the broader literature on object-given/state-given distinction.

4.4 – A Pluralist Proposal: Normativity and Criticizable Failure

I now need to explain how it could be true that someone ought to do something, even if they have no reason to do that thing. In this section, I will provide a proposal which I believe illuminates the plausibility of normative pluralism.

To begin, it is helpful to think, in the most general terms available, about what is involved in it being true that one ought to do something. It is tempting to endorse immediately the reasoning behind the Reasons Challenge and respond that the defining characteristic of it being the case that one ought to do something is that one has decisive reason to do that thing. But that answer is insufficiently general. A more general answer is available, and it is one that does not assume that the only norm-generating standard is the reasons standard.

Here is a proposal: the most general thing we can say about someone who ought to X is that if she does not X then she will have made a real mistake. When you ought to do something, then you must do that thing, not in the sense that you cannot avoid doing it, but in the sense that failing to do it would be a legitimate *failure*—a failure of the sort that would warrant *criticism*. And, continuing to speak in maximally general terms, we can observe that any behavior or action which constitutes a genuine failure is a failure to do something required; if it was not the case that one ought to have done differently than one did, then one's doing as one did would not really be a failure at all. So this is what separates normative

⁵⁸ As I indicated in a footnote in Chapter 3, there is an interesting question about how to categorize a subjective-reasons theory of rationality. It is worth flagging here that there could be a subjective-reasons theory of rationality which would involve an endorsement of normative pluralism. One could hold that the standard of objective reasons and the standard of subjective reasons are home to fundamentally normatively distinct, but equally legitimate, types of 'ought'-claims. A view along these lines is suggested in Schroeder's forthcoming "What Makes Reasons Sufficient?"

principles from other principles. A true ‘ought’-claim—one which binds us, which we are required to follow—is a principle which one can violate only on pain of criticizably failing.⁵⁹

Notice that this proposal makes no reference to reasons.

The proposal I have just offered passes the sniff test. It is quite plausible that a requirement is normative if and only if non-compliance with the requirement is a criticizable failure. Nevertheless, the reasons fundamentalist may insist that my appeal to the notions of criticizability and failure is an unhelpful distraction that does not shed any new light on the nature of normativity. To see why one might have this worry, consider how my proposal would have us assess a typical, run-of-the-mill ‘ought’-claim.

Consider the following statement: *You ought to maintain a healthy body*. Ignoring deviant cases, it is safe to say that this ‘ought’-claim is true. Why? Well, in keeping with what I have just said about the link between criticizable failure and the legitimacy of an ‘ought’-claim, we might say that the ‘ought’-claim is true because if you did not maintain a healthy body (e.g., if you ate poorly and never exercised) then you would be making a criticizable mistake. But an opponent will note that this is an uninformative statement of why you ought to stay healthy. There must be something to say about why you would be criticizable if you failed to maintain your health. To provide a more informative characterization, we might note that staying healthy is crucial to many other valuable endeavors, or that maintaining a healthy

⁵⁹ I say “criticizably failing” instead of simply “failing” because not everything which might be called a “failure” is a failure to do something required. If I miss when I try to throw a ball of paper into the recycling bin, I have failed, but I have not failed in a normatively relevant sense. The notion of criticizability helps to pick out the relevant class of failures. Of course, one might try to do this in other ways. (To name one example, one might propose that normative requirements are principles which it would be an “objective” failure not to satisfy. But this strikes me as a murkier and more problematic proposal than mine.) I will say more about criticizability later in this chapter.

body will help to prevent you from future disease and suffering, or that being healthy feels good. However, in offering this fuller explanation of why not maintaining a healthy body would be a criticizable failure, all we will have done is list your *reasons* for staying healthy. My reasons fundamentalist opponent will point out that you are criticizable for ignoring these reasons because an agent is always criticizable for failing to do what she has decisive reason to do. Looking at this case may make it seem as though appealing to the more general notion of a criticizable failure is pointless. If the class of criticizable failures is identical to the class of failures to do what one's reasons demand, then we might as well forget this extra business about criticizability.

This reasons fundamentalist's objection is just another instance of the Reasons Challenge. But now we can show why it comes up short. Of course, when we are operating within the domain of giving and receiving reasons, we do not need to keep reminding ourselves that failing to respond to one's reasons is a real failure. When we are in the business of offering and assessing reasons, we take for granted the criticizability of non-compliance with our reasons. But should not take this to mean that criticizably failing *is just the same as* failing to do as one's reasons require or that *whenever* an agent criticizably fails it is because she is criticizable for failing to do what she has reason to do. There can be standards other than the reasons standard which it is a failure not to live up to. It is easy to lose sight of this fact, since such a dominant aspect of rational agency, and interpersonal interaction, consists of providing and considering our reasons. But the notion of criticizability can still do quite a lot of work. If there are a variety of dissimilar grounds on which an agent's behavior can constitute a criticizable failure, then there are a number of independent standards according to which legitimate 'ought'-claims can be made.

Recall the discussion Korsgaard's constitutivism in Chapter 2. Korsgaard proposes

that all normative requirements flow from the standards of agency. She argues that because we cannot help but be engaged in the activity of agency, the standards of agency are standards which we are required to live up to. As she puts it, “a constitutive principle for an inescapable activity is unconditionally binding.”⁶⁰ Now, it does not matter too much whether you agree with the details of Korsgaard’s argument. What matters for my purposes is that she highlights a perfectly intelligible way in which an agent could commit a criticizable mistake that is not a mistake of failing to comply with some independent set of reasons. We do not have persistent and decisive reason, in the reasons fundamentalist’s sense, to comply with Korsgaard’s principles, and Korsgaard is not alleging that we do. The mistake made by the Korsgaardian non-self-determiner is of an entirely different variety. You do not need to agree that non-self-determination is, in fact, a criticizable mistake. You need only to accept that there can be a meaningful, substantive discussion about whether this is a form criticizable failure. Accepting even this much is to admit the coherence and possibility of normative pluralism.

When we are assessing the legitimacy of a proposed ‘ought’-claim, there are at least three sequential questions which must be asked. The first question is: what is the standard relative to which the ‘ought’-claim is being made? The second question is: is it the case that not meeting the standard in question would constitute a legitimate failure on the part of the agent? The third is: is it the case that the proscribed action (or inaction), state of mind, way of being, or what have you, is one demanded by the standard in question. Think again of keeping a maintaining a healthy body. To assess the legitimacy of the ‘ought’-claim on offer (*You ought to keep a healthy body*), we must ask the three questions just listed. The answer to the

⁶⁰ Korsgaard 2009b, p. 39.

first question, arguably, is that the ‘ought’-claim is made relative to the reasons standard. In claiming that one ought to keep a healthy body, we are claiming that one’s reasons decisively favor keeping a healthy body. Now we can ask the second question: is it the case that not meeting the standard in question (i.e., the reasons standard) would constitute a legitimate failure on the part of the agent? The answer—which is a substantive answer, requiring normative judgment—is “Yes.” Then, for the third question, we must ask: is it true that the reasons standard demands that we keep a healthy body? Here, too, the answer is “Yes,” which is just to say that we do indeed have (decisive) reason to keep a healthy body. So, in sum, it is true that you ought to keep a healthy body because there is a legitimate standard which you would be failing to live up to if you did not keep a healthy body. The truth of normative pluralism resides in the fact that there is more than one legitimate standard relative to which failure-assessments are made. Not every criticizable failure is a failure to behave in accordance with the demands of one’s reasons.⁶¹

This gives us the resources to respond to the Reasons Challenge and to diagnose why it seemed like such a challenge in the first place. The Reasons Challenge asks us, in short: “How could it be that I ought I to X if I have no *reason* to X?” If in response we attempt to provide some consideration in virtue of which one ought to X, then we will—so

⁶¹ I am in the process of attempting to open the door to the possibility that some legitimate normative claims are made relative to a standard other than the reasons standard. But one might put pressure on a different point. One might suggest that I have been going too easy on the reasons standard itself, and so might ask for clarification about what, exactly, the reasons standard *is* and why (and in what sense) deviation from that standard is a criticizable failure. As with other issues concerning the nature and structure of reasons that I have confined to footnotes, I am afraid that this is an issue that I could not hope to adequately deal with in this project. My view—which I cannot defend here—is that a failing to comply with one’s reasons is a matter of failing to appreciate or respond to the good-making (or truth-making) properties in the world.

says our opponent—be providing a reason for X-ing, and so we will not have met the challenger. Alternatively, if we refuse to answer, then we will be left without grounds for believing that we really ought to X. But now we can see that the puzzle is an illusion. If you ask “Why ought I to X?”, it is a sufficient answer to explain how you would be going wrong if you failed to X. This explanation need not take the form of offering a set of considerations which constitute reasons in favor of X-ing and which you would be ignoring if you did not X. Instead, a full explanation can be given by providing an account of the type of failure that you would be engaging in if you did not X.

I have been arguing at a fairly high level of abstraction. I have been arguing that there is nothing confused in the thought that there may be multiple sources of normative demands and, as such, that there might exist a non-reasons-constituted normative requirement. But I have not yet made progress toward giving an account of a requirement-generating standard other than the reasons standard. And, most importantly, nothing I have said so far bears directly on the questions concerning the requirements of structural rationality. Even if what I have said indicates how we might respond to the general worry created by the Reasons Challenge, I have not say anything about why the requirements of structural rationality are normative. For that, I would need to explain the (non-reasons-based) criticizability of non-compliance with these requirements. I will make progress on this issue in the next two chapters.

Before closing this chapter, I must make a number of important clarifications and respond to a number of crucial worries about the proposal I have offered in this chapter. In the next section, I will march through a number of clarifications concerning the general thesis of normative pluralism, and a number of clarifications concerning the particular pluralist proposal that I am endorsing. Then, in Section 4.7, I will consider a crucial line of

response from the reasons fundamentalist which threatens to undercut the pluralist thesis.

4.5 – Clarifications: Pluralism in General and My Pluralist Proposal

I need to make six points of clarification. The first two points concern the nature of normative pluralism in general. The third through sixth points concern the specific version of normative pluralism that I laid out and defended in the previous section.

First, it is worth emphasizing that one could, in principle, be a normative pluralist without endorsing the proposal concerning criticizable failure that I sketched in the last section. Normative pluralism is the view that there are multiple fundamentally distinct types of ‘ought’-claims, relative to multiple fundamentally distinct normative standards. I have proposed that pluralism is true because (i) an ‘ought’-claim is binding just in case deviation from the claim’s demand is a criticizable failure and (ii) there is more than one fundamentally distinct standard relative to which non-compliance can be a criticizable failure. In making these two claims I am offering a particular pluralist proposal, rather than expanding on pluralism’s definition. I am not alleging that my proposal concerning failure and criticizability is conceptually contained within the bare concept of normative pluralism. I find the criticizable-failure proposal to be the most plausible candidate explanation of pluralism’s truth. But it is important to underline that, in principle, normative pluralism could be defended by a completely different proposal.

Second, the bare pluralist thesis says nothing about how many distinct requirement-generating standards there may be, and it says nothing about what those standards are (however many there are). The thesis of normative pluralism says only that there is more than one such standard. I have accepted that the reasons standard is an ‘ought’-generating

standard and I believe that there are others. But, in principle, one could be a normative pluralist while altogether rejecting the reasons standard as a basic normative standard. There could be a version of pluralism which rejected the notion of a normative reason, or which hold that normative reasons as parasitic on some other, more basic, normative standard. And one could be a pluralist while believing that there are two, or twenty, or two thousand distinct normative standards. As it concerns the version of pluralism that I endorse, I do not have a settled view about the number of distinct grounds of criticizable failure. I believe that there are at least two.

Now onto four important points of clarification about the specific pluralist proposal I offered in Section 4.4. First, you may be wondering whether the pluralism I am endorsing is actually a *pluralist* view at all, given that I am saying that all normative principles share a common property: they make demands on us which warrant criticism when we fail to comply. Might this suggest that the normative domain is unified? In a sense, yes. But not in any interesting sense. The domain of normative requirements is unified only in the sense that all of its members are *normative*. A principle's normativity consists in the fact that deviation from its demands is a relevant kind of failure. But this does not suggest that there is any important relationship between every normative requirement.

Second, I am using “criticizable” here to mean something like “fit to be critically appraised” rather than merely “can be criticized.” Agents and behaviors which are not criticizable *can* be criticized, just like items which are not edible *can* be eaten. So my use of “criticizable” is similar to our uses of “edible”, “wearable,” and “flushable” and unlike our uses of, say, “unbreakable” (since something which is unbreakable really cannot be broken). An ‘ought’-claim is true for an agent when the agent’s deviation from the principle’s demands warrants critical appraisal of the agent.

Third, when I say that an agent is “fit to be criticized” or “fit to be critically appraised,” I do not mean that (necessarily) we should offer outward criticism toward her, or that criticizing her would be a good idea, or even that criticizing her would be permissible. In many cases, criticizing people for their criticizable mistakes is unwise or rude. In part, this is due to the way that we (humans) tend to deal with criticism and due to the likely consequences of criticizing each other. Many of us get upset when we are criticized and have difficulty maintaining healthy relationships with our critics. If we went around criticizing each other every time that one of us made a criticizable mistake, we would never get anything else done. We have good grounds for not criticizing each other in many of the cases in which we are criticizable in the sense that I have in mind. People who drive slightly below the speed limit on open roadways, and people who skip to the end of good books, and people who take professional sports too seriously are all intelligible and proper targets of critical appraisal, and they are all making criticizable mistakes. But it would normally be a bad choice to offer our critical judgments to these people.

Finally I come to the most important point of clarification. One might wonder how it is that we learn whether an agent’s failure to comply with some putative ‘ought’-claim is a failure of the normatively relevant variety. This question gives rise to a set of worries which will be the subject of the next section of this chapter. But as a way of paving the way for that discussion, I offer the following flat-footed thoughts. I said earlier that in assessing whether some putative ‘ought’-claim is legitimate, we must ask (i) what standard is being appealed to, (ii) whether the standard is one which it would be a criticizable mistake to violate, and (iii) whether the proscribed action or behavior is one demanded by the standard in question. The question at hand now concerns the second of these steps: how are we to come to know whether a particular standard issues binding dictates?

There is an easy answer here, and I think that it is the only plausible answer available, but it is one that some will find unsatisfying. The answer is that we must engage in the same style of deliberation and analysis that we engage in whenever we must make a substantive normative judgment. There is no special mechanism or formula for determining whether it would be criticizable to do or be like such-and-such. Which kinds of failures to comply with a demand are criticizable and why they are criticizable are substantive questions the answers to which will require that we attend to and think about the details of the relevant cases and proposals. Consider, again, Korsgaard's constitutivist proposal. To find out whether it is a criticizable failure not to comply with the constitutive standards of action, we have no option other than to consider what this form of putative failure involves and think about why this form of failure would be relevant to us. No more specific method is available. And the fact that there is no more specific method available should not make us doubt that there are facts about when an agent is and is not criticizable for a failure to comply with some proposed demand.

4.6 – A Covert Appeal to Reasons?

In making my sixth point of clarification in the last section, I may have inadvertently raised a red flag. I just proposed that determining which standards are 'ought'-generating standards requires normative judgment, because there is no non-normative mode of evaluation which will inform us whether a failure to be like such-and-such is a criticizable failure.

The difficulty for my proposal—and, specifically, for its status as a pluralist proposal—is that a judgment about whether a particular standard serves as the basis for legitimate 'ought'-claims appears to be a judgment about *reasons*. If we conclude that

deviating from the claims of Standard A is a criticizable mistake, while deviating from the claims of Standard B is not, then what sort of normative judgment are we making? Plausibly, we are judging that we have good grounds for caring about whether we comply with the demands of Standard A and that we do not have such grounds for being concerned about complying with the demands of Standard B. And this is just a judgment about our reasons for caring about compliance with these standards. My account, so goes the charge, involves a covert appeal to reasons in determining the normativity of any ‘ought’-claim. And this might undermine my contention that I am putting forward a form of normative pluralism according to which there can be non-reasons-based normative claims.

I must admit that there is something to this characterization of my view. I will readily admit that judgments about which standards (and so which forms of failure) are relevant are normative judgments. They are judgments about which standards are important to comply with and which are not. And this is very close to saying that they are judgments about our reasons for caring (or not caring) about complying with the standards in question. It approaches unintelligibility to say that we could be bound by a principle or standard which we have no reason to care about complying with. Nevertheless, I do not think that these insights raise a serious problem for my view.

The first question to ask is whether my view, in appealing to reasons-concerning judgments, somehow collapses into a version of standard, reasons-fundamentalism. Here the answer is obviously “No.” As I explained in Chapters 1 and 3, reasons fundamentalists hold that the only normative ‘ought’ is the ‘ought’ of most or decisive reason. On their view, a statement of the form *A ought to X* is true just in case A has decisive reason to X. If the worry expressed in the previous paragraph has teeth, then it may be that I am committed to the view that a claim of the form *A ought to X* is true just in case A has

reason to care about whether she Xs. But having a reason to care about whether one Xs is not the same as having (decisive) reason to X. Indeed, it does not follow from one's having reason to care about X-ing that one has any reason at all to X (let alone *decisive* reason to X). Reasons for caring about compliance are, in many cases anyway, distinct from reasons for compliance. So even if a judgment about which standards are binding is (always?) a reasons-involving judgment, it does not follow that every requirement's normativity is constituted by reasons for complying with that requirement.

A critic may claim that reasons are still featuring too prominently in this allegedly pluralist account. If a principle is normative only if it is one we have reason to care about complying with, then it is still reasons that are doing all of the important normative work. It may seem as though reasons, or reasons-related facts, are still what make a principle normative, because it is in virtue of these reasons (to care about compliance) that a standard can generate binding requirements. If these reasons did not exist, then the standard would and could not be requirement-generating.

This worry is misguided. I can accept that a principle is normative only if it is one that we have reason to care about complying with. But the fact that we have reason to care about complying with the principle is not a part of the explanation of the principle's bindingness. The reasons we have for caring about compliance are produced by the ultimate source of the principle's bindingness, but are not themselves that source. Take some putative non-reasons-constituted 'ought'-claim demanding that you X. In order to assess whether the 'ought'-claim in question is binding, we must ask whether it would be a mistake to fail to live up to the X-demanding standard. Suppose that the answer is "Yes." If so, this is due to certain facts about what is involved, for a being like you, in not living up to the standard which demands that you X. These facts, whatever they are, will at the same time explain why

your failure is criticizable *and* why you have reason to care about complying with the requirement. This is simply because any considerations which make non-compliance a criticizable failure will also deliver reasons for caring about whether you comply. This should be no surprise. If your not-X-ing is something that would be a failure on your part of the sort that would warrant critical appraisal, then of course you will have reason to care about whether you comply with the demand. But the reasons for caring are not doing the work of making it the case that you ought to comply. You ought to comply just because your non-compliance would be a mistake in some (here unspecified) way. So even though judgments about which standards are binding involve judgments about our reasons for caring about compliance, facts about our reasons for caring about compliance are not themselves what make any requirement normative.

There still remains an interesting question regarding whether there could be a completely reasons-independent form of normative ‘ought’-claim. For example, we might wonder whether there could be an ‘ought’-claim that bound a non-reasons-responsive intender and believer. Or, a bit differently, we might wonder whether in a ‘reasonsless’ world (if there could be such a thing), it would still be true of beings like us that there were things that we ought to do or ways that we ought to be. I suspect that the answer is “No”: if we had no reasons, or if we were non-reasons-responsive beings, then there could be no normativity at all. This suggests that reasons have a special place in the world’s normative architecture, but it does not suggest that all normative claims are claims about our reasons. I will not further pursue these issues here.

To close this section, and the chapter, I will make one final point of clarification. There is a distinction that I have so far been intentionally blurring because it does not affect any of my core claims. There is a difference between the claim that a principle’s normativity

is constituted by reasons—in other words, that the principle is normative *because* it is one we have reason to comply with—and the claim that all normative principles are principles with which we have reason to comply. The first claim is much stronger than the second. One could hold that any normative principle or claim is one with which we have reason to comply without holding that every normative principle's normativity is delivered by the reasons we have for complying. For example, the reasons we have for complying with some normative principles could be incidental to, or in addition to, the principles' non-reasons-constituted normativity. Although I have on occasion rather loosely said that there can be normative requirements with which we have no reason to comply, the important point is only that there can be normative requirements whose normativity is not constituted by reasons.

I am open-minded as to whether we do have some reason, of some strength, to comply with some or all of the non-reasons-constituted normative principles which bind us. For example, it may be that because failing to comply with a legitimate demand is a failure, and because we have reason not to make criticizable mistakes, we always have some derivative reason to bring ourselves into conformity with the normative requirements to which we are subject. I have no settled judgment about this question. If this turns out to be true, then it does no harm to the pluralist thesis, since this would not suggest that the normativity of every legitimate principle is just the normativity of reasons. It would still be true that there was more than one legitimate sense of 'ought' and more than one legitimate type of 'ought'-claim.

4.7 – Next Steps

My defense of normative pluralism in this chapter has been merely provisional, because I

have come up short of establishing that there are any requirements whose normativity is not constituted by reasons. I have not yet show that there are any requirements it is a criticizable failure to violate regardless of whether we have any reasons for compliance. And, most importantly, I have said nothing about the normativity of the requirements of structural rationality.

In isolation, my arguments in this chapter have established relatively little. But this ground-clearing work will prove critical to my arguments in the rest of this project. In Chapters 5 and 6, I will concretize and put to work the pluralist proposal I have just finished detailing. In the next chapter, I will uncover the standard that undergirds all of structural rationality. Doing so will solve the content problem of rationality, while at the same time providing us with an explanation of rationality's normativity. As we will see, once we have understood what rationality, as a whole, is 'about'—once we have revealed the unifying standard of structural rationality—there will be no deep question concerning why we are criticizably failing when we fail to meet its demands.

5.1 – A General Requirement of Structural Rationality

My positive argument has two crucial pieces, one of which is now in place. In Chapter 4, I argued that a putative requirement is binding for an agent—that it is one with which she ought to comply—as long as her violation of the requirement would be a criticizable failure. Now, in this chapter and the following chapter, I will show that the requirements of structural rationality are principles of this sort. Once both these steps are secured, I will have provided a realist account of structural rationality which explains the normativity of its demands.

In this chapter, I will not yet attend directly to the question of why we are criticizably failing whenever we violate a rational requirement. Instead, I will finally look directly to the requirements of rationality themselves, in an effort to uncover what rationality, as a whole, is ‘about.’ In Chapter 1, I introduced the content problem: the problem of drawing the boundaries of rationality and explaining what makes requirements of rationality co-members of a single domain of principles. A plausible theory of rationality must be able to explain why principles like IC, BC, ME, and EI are categorically related normative requirements, whereas some other normative requirements (like, for example, a requirement to follow just laws or a requirement to aid the needy) are unrelated. We need a theory that expresses and explains the unity and content of rationality. In this chapter, I will endeavor to provide such a theory.

My primary conclusion in this chapter will be that there is an ultimate requirement of structural rationality from which all the familiar requirements of rationality can be derived. Although the domain of rationality appears to be populated by requirements concerning consistency in intention and belief, means-end coherence, enkrasia, and so forth, these

requirements have a shared foundation in a fundamental requirement which I will call *the General Requirement of Structural Rationality* (or “General Requirement” or “GR”). The formulation I favor, stated in the broadest terms available, is this:

The General Requirement of Structural Rationality
You are rationally required not to attitudinally undermine the functional success of your agential attitudes.⁶²

The General Requirement prohibits attitudinal self-undermining or, in other words, it demands that an agent not ensure that some of her attitudes will fail directly as a result of how those attitudes are employed.

Introduced so abruptly, it is bound to be unclear why this is the most plausible expression of rationality’s ultimate demand. In the course of this chapter I will construct an “unpacked” formulation of the General Requirement which lays bare its contours and implications, and I will argue that every other requirement of structural rationality can be derived from it. If I am correct, then violations of the various well-known requirements of rationality are in fact simply violations of the General Requirement. In other words, every irrational agent is, at root, exhibiting precisely the same attitudinal mistake.

The General Requirement is, by itself, enough to solve the content problem: it tells us what rationality as a whole is ‘about,’ and it gives us guidance in articulating the members of the rational domain. Solving the content problem is a significant achievement in its own right, but it is not the only benefit of finding the General Requirement. In solving the content problem and explaining the unity of rationality, the General Requirement illuminates

⁶² Throughout Chapters 5 and 6, whenever I say “attitudes” or “agential attitudes,” I have in mind only intentions and beliefs. There are interesting questions about how and whether rationality constrains our other kinds of attitudes (e.g., suspicions, desires, emotions), but I cannot attend to those questions here.

the underlying standard relative to which every ‘ought’-claim of rationality is made. The General Requirement shows us what every irrational agent is doing (or failing to do) and so shows us what standard irrational agents are failing to live up to. As we will see, this will give us insight into why irrationality is a criticizable failure. By solving the content problem, and seeing what rationality is ‘about,’ we will begin to see why its claims are normative. But, as I said, I will not here go all the way toward conclusively establishing that violating the General Requirement is criticizable—that will be the business of Chapter 6.

5.2 – Methodology

Before beginning, I need to say something about the process that I will be relying on. To arrive at the correct formulation of the General Requirement, all that we can do, I believe, is propose possible formulations of GR, test those formulations against salient and intuitively clear cases, and assess whether the formulations’ verdicts about such cases are acceptable. This means that searching for a proper formulation of the General Requirement will depend partially on judgments about rationality that we bring to the investigation. In some cases, we must reject a formulation of GR because it produces what seems to be the wrong result—because it declares that an agent is suffering from structural irrationality when that seems to not be the case, or because it cannot find the fault in an agent who seems like a paradigmatic example of a structurally irrational agent. But, in other cases, we might reasonably allow our less firmly held judgments about certain cases to be revised or abandoned, if an otherwise-appealing formulation of GR produces an unexpected verdict about a case.

Of course, I am describing a method of *reflective equilibrium*. But although this methodology is a commonly employed procedure in other areas of philosophy—and

especially normative philosophy—some might find it to be a worrisome practice when thinking about the requirements of structural rationality. One might claim that our starting-point judgments are sparser, less precise, and less firmly held in the case of structural rationality than in the case of, say, morality, and that for this reason we ought not to feel entitled to carry out the investigation in the same way.

I am confident that the methodology is a justifiable one and that, in fact, the similarity between investigations of structural rationality and morality are quite strong. There are a number of requirements of structural rationality which we can use as fixed points in an investigation. Any non-skeptical investigation into structural rationality must assume that the requirements set out in Chapter 1—BC, IC, ME, EI—are legitimate requirements in need of explanation. Versions of these requirements can be used to test proposed formulations of the General Requirement; if a formulation cannot derive a plausible version of these requirements, then it must not be an appropriate formulation.⁶³ At the same time, we can accept that some other accepted judgments about structural rationality are less secure and may reasonably be revised or abandoned for good theoretical reasons. This is similar to a moral philosopher holding that her moral theory must be able to explain the wrongness of promise-breaking, while simultaneously (I) acknowledging that she is not (yet) confident about precisely what form the prohibition on promise-breaking must take, and (II) also being willing to revise or abandon certain pre-theoretical moral judgments that are less entrenched than her judgments about the wrongness of promise-breaking.

With these thoughts in mind, I will now endeavor to build a formulation of the

⁶³ One would then be left to judge whether (i) this is because there is not a single general requirement from which all such requirements can be derived or (ii) this is because the proposed formulation of the General Requirement is just not the proper formulation.

General Requirement that illuminates its details. In Sections 5.3 and 5.4, I will focus on some of the most familiar principles of structural rationality to help me reach a plausible formulation. In Sections 5.5 and 5.6, I will investigate what my account has to say about some less-familiar principles.

5.3 – Consistency in Belief and Intention

The proposal I favor is undergirded by a certain conception of intention and belief. Specifically, my proposal assumes that beliefs and intentions each have a particular constitutive function. Although the conception of intention and belief I have in mind is not immune to criticism, it is broad enough and general enough that it should not be the source of much controversy.

We can illuminate the functions of belief and intention by reflecting on what an agent *is doing* when she believes P and, alternatively, when she intends to X. Consider the case of belief. What is a person “up to,” as it were, when she has a belief? A person who believes P is taking a particular proposition to be true or, put differently, taking a certain state of affairs to obtain. For instance, if you believe that it is raining outside, we might describe your mental state by saying that you have an is-true attitude toward the specified proposition. Conceived of in this way, the function of a belief is to represent (some part of) the way things are. Accordingly, we might then say that your belief succeeds, functionally speaking, just in case it is true. A related point can be made about intending. A person who intends to X is setting out to make a proposition true or, put another way, setting out to bring about a specified state of affairs. If you intend to open your umbrella above your head, you are setting out to make it obtain that your umbrella is opened above your head. While

the function of a belief is to represent the way things are, the function of an intention is to change the way things are. An intention succeeds, functionally speaking, just in case it makes its object true.⁶⁴

I am not proposing a full theory of belief and intention, and I have no need for one. I am only characterizing a general way of thinking about what intentions and beliefs are. In short, I am proposing that a belief is an is-true attitude and an intention is a make-true attitude. According to this way of thinking, a belief succeeds when its object is true, while an intention succeeds when it makes its object true. Although I submit that these ideas are extremely plausible, a defense would require a significant detour. I believe that it would be a mistake to venture into that territory here.⁶⁵

If these claims are granted, we can observe that many cases of structural irrationality have something interesting in common: they involve agents whose attitudes cannot all succeed. An agent who is violating one of the well-known requirements governing

⁶⁴ For the remainder of this chapter, when I talk about the success or failure of an attitude, I always have in mind *functional* success or failure. It is plausible that there are other standards relative to which beliefs and intentions can succeed and fail. For example, a false belief that is adopted on the basis of overwhelmingly strong evidence could be called “successful” in a certain sense. Even more naturally, one might say that it is a kind of “good belief.” But that is not the standard I have in mind. It does no harm to my view that there may be other legitimate standards by which beliefs and intentions may be assessed.

⁶⁵ While many people find this view about the functions of intention and belief plausible (or even obvious), I admit that some may not. As I have tried to stress above, it is not—and cannot be—a primary focus of this particular chapter to defend the view, and so I must take it as an assumption. However, I believe that the arguments in the remainder of this chapter offer defeasible evidence in favor of this assumption about the functions of belief and intention: in short, the fact that this view about belief and intention can do so much explanatory work is reason to take it seriously.

consistency of intention or belief will be such that at least one of her beliefs or intentions must fail. Consider the familiar requirement of belief consistency which I have previously labeled “BC”:

BC

You are rationally required not to [believe P, believe $P \rightarrow Q$, and believe $\neg Q$].

When you violate BC your beliefs cannot all be true. Now I can introduce a related and similarly familiar requirement prohibiting directly contradictory beliefs.

BC2

You are rationally required not to [believe P and believe $\neg P$].

If you violate BC2, it must be that at least one of your attitudes fails.

These rather mundane and obvious facts are enough to deliver a proposal. Perhaps what unifies all of structural rationality is the basic rational impermissibility of having a set of attitudes in which at least one of the involved attitudes must fail. We can use the label “GR1” to refer to this provisional formulation of the General Requirement.⁶⁶

⁶⁶ Proposals along these lines have been considered but quickly rejected by several writers. For instance, on p. 151 of Broome 2013, Broome writes: “[A]s people say, it is in the nature of beliefs that they aim at truth. Since two contradictory beliefs cannot both be true, that may explain why it is irrational to have contradictory beliefs. This seems plausible, but I do not know how this general idea can be worked out in detail, to provide a criterion for determining what rationality requires.” In Kolodny 2008a, Kolodny notes in passing that when one has a pair of intentions which one believes are incompatible at least one of one’s attitudes will fail. Kolodny does not pursue the underlying thought that all of structural rationality is rooted in a more general prohibition on attitudinal self-undermining. Raz 2005a discusses the putative irrationality of harboring a set of inconsistent beliefs, but concludes that the guarantee of having a false belief is insufficient to serve as the basis of a general account of irrationality. These writers would be right to reject an account which presents GR1 as the unifying principle of structural rationality. Nevertheless, as I will explain, there is a much more promising account nearby.

GR1

You are rationally required not to have a set of attitudes which is such that at least one of the involved attitudes must fail.

One promising feature of GR1 is that it shows that practical and theoretical requirements can have a common source. As we have seen, GR1 can be used to quickly derive two theoretical requirements. It can also be used to derive a version of a familiar practical requirement that I have been referring to throughout this project. Recall the requirement of intention consistency which I have been calling “IC”:

IC

You are rationally required not to [intend to X, intend to Y, and believe that you cannot both X and Y].

If you violate this principle, you do so by intending to X, intending to Y, and believing that you cannot both X and Y. When you do this, at least one of your attitudes must fail. If your belief is true (and so the two intentions are incompatible) then at least one of your intentions is sure to fail. And if both of your two intentions could succeed (together) in realizing their objects, then it must be that your belief is false. In violating IC you have your attitudes arranged such that at least one of them must fail. So IC can be derived from GR1.

This is a good start, but it cannot be the whole story. GR1 offers an incomplete and potentially misleading characterization of what is distinctive about cases of structural irrationality. To begin working toward an improved formulation of the General Requirement, here is a small but important point of clarification. Consider the “must” invoked in GR1. Whether GR1 is plausible depends crucially on the modality of the term. Imagine, for example, that you both intend to drive your mother to the airport tonight and intend to bake a cake tonight. Although you have not thought about your plans in great detail, you imagine that you will spend a couple of hours in the car with your mother and then return home to begin baking the cake. But you have badly miscalculated. You have

forgotten that driving your mother to the airport is actually a huge undertaking, since the airport is at least four hours away by car. By the time you return home from the eight-hour round trip, it will be long after midnight, and baking a cake during the (previous) evening will no longer be an option. None of this is at all apparent to you, and so you continue to both intend to bake a cake tonight and intend to drive your mother to the airport tonight. What are we supposed to say about your rationality in this case? Intuitively, it would be a mistake to deem your behavior in violation of any requirement of structural rationality. Someone in your position has an awful plan, but uncritically planning to do more than you could actually accomplish is not the sort of thing that can render you irrational. Nevertheless, there is a clear sense in which one of your two attitudes really *must* fail. Accomplishing both the mother-driving and the cake-baking in the same evening is not an open possibility for you. So it looks as though GR1 may condemn you for having the relevant arrangement of attitudes. This is the wrong result.

The obvious response is that the “must” contained in a proper formulation of the General Requirement invokes a broader form of possibility than is involved in our judgment that it is “impossible” for you to accomplish both your cake-baking and mother-driving tonight. After all, the world could easily be a bit different—your car could have been built to move faster, or the airport could have been stationed closer to your home, or you could have learned to bake a cake while driving—and then your two intentions could have jointly succeeded. In the relevant sense of “must,” it is not the case that your attitudes are structured so that at least one of them must fail.

By invoking a broader form of possibility, GR1 can be revised to accommodate this small point. Instead of GR1, we can now consider

GR2

You are rationally required not to have a set of attitudes which is such that,

logically, at least one of the involved attitudes must fail.

This is a small modification, but it makes an important difference. We can still use GR2 to derive the three genuine requirements listed above. At the same time, GR2 gets the appropriate result in the mother-driving case, by not rationally condemning you when you intend to drive your mother to the airport tonight and intend to bake a cake tonight. GR2 helps to clarify an ambiguity in GR1. And this is an important lesson to be carried through the rest of this investigation: the attitudinal failure involved in structural irrationality is *logically* guaranteed.

GR2 is an improvement, but it, too, cannot be a proper formulation of the General Requirement. To see why, consider the following case. Imagine that after reading a (vandalized) Wikipedia entry on prime numbers, you come to believe that Euclid's argument has been disproven and that there are actually only finitely many prime numbers. Of course, your belief is false. And, presumably, your belief is *necessarily* false—it is a logical impossibility that there are only finitely many prime numbers. Call your belief *b*. Any set of attitudes containing *b* as a member (including the set of attitudes containing only *b*) will have at least one attitude (namely, *b*) which is, logically speaking, guaranteed to fail. This means that you will be in violation of GR2, since GR2 prohibits sets of attitudes in which at least one member is bound to fail, logically speaking. But this is an odd result. You are not irrational for having the false belief about prime numbers. And there is nothing special about the Euclid case; in general, one is not irrational simply for having a belief in a necessarily false proposition.

The problem with GR2 is that it does not capture the fact that structural irrationality is a *structural* matter. GR2 suggests that the problem with the structurally irrational agent is just that, within a given attitudinal set, it is impossible for all of her attitudes to succeed. This

is a misleading characterization. It is true that the structurally irrational agent is bound to have at least one failing attitude. But that is not the most important point. The most important point is that some of her attitudes are bound to fail because of the way her attitudes are arranged or combined. The structurally irrational agent has a combination of attitudes which is such that it is logically impossible for all of the attitudes to succeed, due to the relations between those attitudes. We can revise GR2 to be responsive to these considerations. Consider:

GR3

You are rationally required not to have a set of attitudes which is such that, logically, due to the relations between the attitudes in the set, at least one of the involved attitudes must fail.

In Euclid-type cases, you are not violating GR3. Although you have a belief that is guaranteed to fail, the failure is not due to the relations between your attitudes, and so you are in compliance with GR3's demands. GR3 is a clear improvement on GR2. It allows us to derive all of the genuine requirements of structural rationality listed above and it avoids getting the wrong result in Euclid-type cases. At the same time, GR3 begins to shed some light on the fundamental standard that undergirds rationality: a prohibition on *attitudinal self-undermining*. So GR3 brings us much closer to the truth. But it also generates verdicts that some may find unsettling. It is worth exploring some of these implications.

If GR3 is correct, then there are no particular attitudes which are, in themselves, banned by structural rationality. Because GR3 prohibits only sets of attitudes containing members that stand in certain relations to each other, GR3 can only prohibit sets of attitudes containing at least two members. On its face, this may not look like a problem. After all, the tendency for GR2 to prohibit single-member sets of attitudes was partially responsible for generating the problematic conclusion in the Euclid case. And, in the end, I am quite happy with GR3's blindness to single attitudes. But considering certain cases may make one worry

that GR3 has taken us off track.

For example, consider an agent, Mark, who has the following set of two beliefs: (i) *that O is a perfect circle* and (ii) *that O is not a perfect circle*. Mark is violating BC2, according to which you are rationally required not to [believe P and believe \neg P]. This requirement is immediately derivable from GR3, just as it was derivable from GR2 and GR1. There is nothing weird about this since Mark does indeed seem to be exhibiting a standard form of structural irrationality in belief.

Now consider a different agent, Luke, who has a single, conjunctive belief *that O is a perfect circle and O is not a perfect circle*. GR3 has nothing to say about Luke's rationality, because the guaranteed failure of Luke's attitude is simply the result of the content of that attitude, rather than the result of the relations between any attitudes. Just as GR3 was silent on your rationality in the Euclid case, GR3 is silent on Luke's rationality. This might strike you as a bad result. You might think that if Mark is violating a principle of rationality, then so must be Luke. And you might think that if GR3 cannot accommodate this point—by reaching identical verdicts about Mark and Luke—then GR3 must not be a legitimate formulation of the General Requirement.

I think that GR3 can be retained despite this seemingly odd result. In any case that suitably mirrors the Mark/Luke example, an agent with a Luke-type conjunctive belief will in fact be violating GR3, because he will, in addition to having a single conjunctive belief, have independent beliefs in each of the conjuncts. In the case at hand, it is not plausible that Luke would believe *that O is a perfect circle and O is not a perfect circle* but not believe both (i) *that O is a perfect circle* and (ii) *that O is not a perfect circle*. I am not making a point about rationality here. I am not saying that in virtue of believing the conjunction Luke is rationally required to believe each of the conjuncts. (If I said that, then I would have to explain the basis of that rational

requirement.) Instead, I am saying that a belief in a conjunction just *does not come* without a belief in each of the conjuncts. This is a fact about the constitution of our mental lives, and the nature of belief, rather than a normative fact. If I am correct—if Luke has these two independent beliefs—then Luke is violating BC2 and, so, GR3 as well. Our inclination to deem Luke irrational is justified by the fact that he is actually in Mark’s doxastic position.⁶⁷

One might argue that it is not strictly speaking *impossible* for an agent to believe a conjunction without believing each of its conjuncts. However, I think we must accept that someone in such a position would not necessarily be irrational. Such an agent would believe something that is, as it so happens, necessarily false. But as is suggested by our judgment in the Euclid case, there is nothing immediately irrational about believing a logical impossibility. Taking this stance creates pressure to explain what it is about “splitting up” the belief into multiple beliefs that generates a violation of structural rationality. In the final section of this chapter, in exploring the normativity of structural rationality, I hope to shed light on this question. For now I must set it aside. The correct move is to retain GR3.

The challenge just considered proposes that GR3 fails to capture certain cases of irrationality. GR3 faces challenges in the opposite direction, too. It may seem that GR3 prohibits certain combinations of attitudes that we would not traditionally judge to be structurally irrational. One obvious case is worth focusing on.

Up to this point, I have presented IC as saying that you are rationally required not to [intend to X, intend to Y, and believe that you cannot both X and Y]. Certainly, GR3 allows us to derive IC. But GR3 also gives rise to a broader requirement concerning the consistency

⁶⁷ Perhaps not identical. Mark may lack a belief in the conjunction. Although a belief in a conjunction is always accompanied by a belief in each of the conjuncts, an agent can believe P and believe Q without also believing [P & Q].

of our intentions. Specifically, from GR3 we can conclude that you are rationally required not to [intend to X and intend to not X]. An agent with such a combination of intentions has, due to the relations between her attitudes, guaranteed that at least one of those intentions will fail. So GR3 finds rational fault in an agent who has two intentions with directly incompatible contents, regardless of whether the agent has any belief about the incompatibility of the intentions. This is likely to be a surprising result, because most paradigmatic cases of irrationality through intention inconsistency involve a crucial element of belief.⁶⁸

I do not find this implication of GR3 unpalatable. Imagine that I have two intentions: I intend to touch the button with my finger at 1:00pm sharp and I intend to not touch the button with my finger at 1:00pm sharp. I am inclined to view this case as a straightforward instance of structural irrationality. A belief does not need to be in the picture to generate structural irrationality in intention. A belief is normally present in these cases, because we rarely have intentions with directly incompatible contents. In a normal case, I intend to X and intend to Y, and it is only in also believing that I cannot both X and Y that I come to have a set of attitudes which is mutually unsatisfiable. In these standard cases, a belief about the (perceived, practical) impossibility of realizing two different intentions is what generates the (actual, logical) structural inconsistency of the three involved attitudes. But a belief is not strictly required in order to create a case of structural irrationality in intention. So I think that we should accept:

⁶⁸ See, for example, Scanlon 2007 and Kolodny 2008a.

IC2

You are rationally required not to [intend to X and intend to not X].⁶⁹

To close this section, I will briefly focus on a rather different sort of reason for doubting that GR3 is a proper formulation of the General Requirement. Some philosophers are dialetheists: they believe that a proposition can be true while its negation is also true. This raises an interesting question for GR3 (and for any account of rationality). The interesting question is not what dialetheists might say about GR3. Instead, the interesting question is what those of us investigating structural rationality should say about the dialetheists. Suppose a committed dialetheist both believes P and believes \neg P. Suppose also that, although the dialetheist acknowledges that these are contradictory beliefs, she maintains her beliefs on the grounds that the beliefs are both *true*. According to GR3 and BC2, she is structurally irrational. But we might worry that this is an unfair result. She is (by stipulation) an expert in logic with a reasoned justification for her position. We might think that, regardless of whether her peculiar view ultimately succeeds, it would be a mistake to endorse a principle of structural rationality which automatically charges her with irrationality. Perhaps we should treat her case as different, and modify GR3 to allow for such exceptions.⁷⁰

I do not think that we should modify GR3. My response to this concern is in part to

⁶⁹ A more precise formulation of IC2 would say that you are rationally required not to [intend to X-*at-t* and intend to not X-*at-t*]. The formulation of IC2 provided in the body of the chapter could be interpreted as prohibiting you from, for example, intending to go to the gym and intending to not go to the gym, even if those intentions were held with respect to different times. That would be a mistake, and so a precise formulation of IC2 would not have this implication. As I noted in Chapter 1, I am not here focused on the inner details of the lower-level requirements of structural rationality, and so I will continue to ignore these complexities.

⁷⁰ See Broome 2013, p. 91 and p. 155.

dig in my heels and in part to suggest that the problem is less serious than it appears.

Assuming that dialetheism is false, it is simply true that an agent who believes P and believes \neg P is violating the demands of structural rationality, even if she has putative justifications for their beliefs. That is the heel-digging part. But (here is the second part) it is important to note that GR3 does *not* imply that endorsing the dialetheist's axiom is irrational. If the dialetheist is violating a requirement of structural rationality, she does so when she believes P and believes \neg P, but not when she makes the (higher-order) judgment that there can be a proposition which is true while its negation is true. It may be odd if GR3 committed me to saying that Graham Priest is structurally irrational just in virtue of his views about logic. But endorsing GR3 does not commit me to saying this.

5.4 – Means-End Coherence

I began with the idea that it is structurally irrational to have a set of attitudes in which at least one of the involved attitudes must fail. It turned out that this was too broad a characterization of structural irrationality. An agent is only structurally irrational if at least one of her attitudes will fail *because of the way that they are structured*. So I revised the working requirement to prohibit only sets of attitudes in which the guaranteed functional failure of an attitude was the result of the relations between the present attitudes.

Still, this cannot be the whole story. GR3 misses out on one of the two key ways in which an agent's attitudinal organization can ensure that at least one of her attitudes will functionally fail. In some cases, as we have seen, the failure is guaranteed because the contents of the present attitudes are related in ways that make their joint success impossible. But in other cases, which I have so far in this essay ignored, the failure is guaranteed by the *absence* of an attitude.

We can see this by turning to requirements of means-end coherence. The finer details of instrumental requirements need not concern us here. A simplified version of central instrumental requirement looks like this:

ME

You are rationally required not to [intend to X, believe that you can X only if you intend to Y, and not intend to Y].

Imagine that you intend to rob the bank at noon, believe that you can succeed in robbing the bank at noon only if you intend to first bail your brother out of jail, and that you lack the intention concerning your brother. In this case, there are two relevant attitudes present: the end-focused intention and the belief about what you take to be the necessary means to your end. You are violating ME and we need to explain why this is so. But GR3 does not give us the explanatory tools.⁷¹

Consider what is going wrong in your case. As long as you do not intend to bail your brother out of jail, at least one of the two attitudes that you *do* have will fail. The possibility of their co-success depends on the presence of the means-focused intention. If your belief is true, and so the success of your end-focused intention does depend on the presence of the means-focused intention, then it is guaranteed that in the absence of the means-focused intention your end-focused intention will fail. Alternatively, if your end-focused intention can succeed even without the presence of the means-focused intention, then this could only

⁷¹ So-called *cognitivists* have argued that ME (and perhaps other intention-involving requirements) can be derived from a theoretical requirement like BC. The standard cognitivist arguments require an assumption that when one has an intention to X, one necessarily has some belief about one's X-ing. (Depending on the version of cognitivism, what belief one is purported to have varies.) These assumptions are contentious. However, the more pressing problem for cognitivists is that they leave unexplained the basis of theoretical requirements of structural rationality. See Setiya 2007 and Wallace 2001.

be because your belief is false. So your end-focused intention and your belief about the necessary means can co-succeed only if you have the means-focused intention.

It is clear that your structural failure is similar to the structural failure prohibited in GR3. As stated, however, GR3 cannot account for your irrationality. In your case, it is not the co-existence of attitudes that guarantees an attitude's failure, but instead the *absence* of an attitude which is required to make possible the co-success of your other attitudes. We can revise GR3 to accommodate the fact that an agent can attitudinally undermine the success of her attitudes in more than one way. Consider:

GR4

You are rationally required not to have a set of attitudes which is such that, logically, due to the relations between its members or due to the absence of some other attitude(s), at least one of the attitudes in that set must fail.

GR4 allows us to derive each of the requirements that I endorsed in Section 3, but also to derive ME.

One interesting though not undesirable implication of GR4 is that it makes it possible for an agent to exhibit structural irrationality in virtue of holding only a single attitude. Articulating this point requires some care. When assessing GR3, I said that the requirement could not prohibit any particular attitude. This remains true of GR4: no single attitude, regardless of its content, is *by itself* rationally off limits. But GR4 could find fault in an agent who holds only a single attitude, if that attitude's functional success depends crucially on the presence of another attitude. This will not change our verdict about cases like the Euclid example. In that case, you have a single belief in a logically impossible proposition. I claimed that you are not irrational for holding that belief. And GR4 would not deem you irrational, because your arrangement of attitudes is not responsible for your belief's falsity. There is no additional attitude that could make successful your necessarily false belief, and so your belief will fail regardless of what attitudes you happen to hold. GR4

is silent on that case.

But imagine a case in which you believe that you intend to donate to charity while, in fact, you do not intend to donate to charity. You are just a miser with delusional beliefs about your own generosity. You are violating GR4, because you have a belief which, due to the absence of an attitude (an intention) from your total network of attitudes, cannot succeed. This is a plausible result, as having these kinds of false beliefs about one's attitudes certainly has the feeling of irrationality. We can accept the following rational prohibition concerning believed intentions:

BI

You are rationally required not to [[believe that you intend to X] and [not intend to X]].

Having the belief while lacking the intention is irrational. There are other related instances of irrationality concerning the relationship between our attitudes and our attitudes about our attitudes. GR4 allows us to account for such forms of structural irrationality.

One might be worried that the move from GR3 to GR4 is ad hoc. All I have done, after all, is insert a clause that allows us to capture cases of instrumental irrationality. But this is the wrong way to look at things. Recall that what caused us to abandon GR2 was its blindness to the fact that the guaranteed functional failure involved in structural irrationality is the result of the way that one's attitudes are structured. In transitioning to GR3, we were focused on a certain range of cases in which the simultaneous holding of certain attitudes makes the co-success of those attitudes logically impossible. This is one way, but only one way, in which an agent's attitudes can have their functional success undermined by the network of attitudes in which they are involved. An agent can also have her attitudes' functional success compromised by 'holes' within her system of attitudes. In some cases, the holding of certain attitudes in the absence of other attitudes makes impossible the success of

the attitudes which are held. The General Requirement of Structural Rationality must be able to accommodate both phenomena. So transitioning from GR3 to GR4 is an effort to accomplish what we set out to do when we revised GR2.

We have arrived at a powerful formulation of the General Requirement. From GR4 we can derive the BC, IC, ME, and a number of other less-often-considered (but intuitively plausible) requirements. GR4 allows us to do all of this without overreaching. I believe that GR4 should be accepted, provisionally, as an accurate formulation of the General Requirement. As I claimed at the outset, the General Requirement of Structural Rationality, concisely stated, says that you are rationally required not to attitudinally undermine the functional success of your agential attitudes. At the time, you might have wondered: how does one attitudinally undermine the functional success of one's agential attitudes? GR4 answers the question: by having one's attitudes structured such that, as a result of the relations between one's present attitudes, or as a result of the absence of an attitude or attitudes, one makes impossible the joint functional success of one's attitudes.

I now want to briefly consider two further classes of structural requirements. In Section 5.5, I will discuss whether there are any closure requirements governing belief. In Section 5.6, I will discuss requirements governing enkrasia.

5.5 – Are There Any Closure Requirements?

Some writers have proposed that not only does rationality require that we not have inconsistent beliefs, and that we not have contradictory beliefs, but also that our network of beliefs is “closed.” Up to this point, I have not considered whether there are any rational requirements governing doxastic closure.

There are various proposals available. According to a very strong proposal, you are

required to believe every proposition entailed by any propositions which you believe. For example, this proposal says:

Where Q is entailed by P, you are rationally required not to [believe P and not believe Q].

GR4 could not be used to derive this principle. According to a somewhat weaker proposal, you are rationally required to believe whatever propositions you *believe* are implied by your beliefs. For example, this proposal says:

You are rationally required not to [believe P, believe that P entails Q, and not believe Q].

Again, GR4 could not give rise to any such principle.

At first glance, GR4's inability to account for requirements of doxastic closure may seem like a problem for my proposal. However, I think that this is the appropriate result. I am skeptical that there are any rational requirements governing closure *per se*. Requirements concerning consistency in intention and belief seem absolutely indispensable to a theory of rationality, as they are the principles which provide our initial conceptual foothold in the domain. But closure requirements do not seem as theoretically attractive or important. Consider the following two points.

First, closure requirements raise intuitive concerns that are not raised by the requirements I have already discussed. Broome and others have pointed out that standard closure principles are exceptionally demanding—far too demanding for any actual agent to satisfy. While perfect compliance with principles like BC, IC, and ME is difficult for imperfect agents like us, compliance with closure requirements is completely out of reach. We simply do not have the cognitive capacity required to satisfy even the weaker of the two closure principles introduced above. So if we are disinclined to see compliance with rationality's demands as unattainable, then it would be a mistake to include closure

requirements in our theory of rationality.

Second, the claim that we are required to have a closed system of beliefs is intuitively under-motivated. As I indicated at the outside of this project in Chapter 1, through the examples of (1), (2), and (3), violating requirements like BC, IC, and ME always seems to involve *getting something wrong*. Structurally irrational agents are agents who are making a mistake, and I submit that it is this prospect of failure that motivates our pre-theoretical judgments about the legitimacy of these requirements. But an agent lacking a closed system of beliefs need not be making any mistake at all. Many of the propositions entailed by (or believed to be entailed by) our beliefs are irrelevant or obscure and, intuitively, a failure to hold these beliefs seems rationally unassailable. At most, we might say that an agent without a closed set of beliefs is not to be taking every possible opportunity to do something (epistemically) good. Accordingly, one could plausibly hold that having a closed set of beliefs is an epistemically laudable although supererogatory goal. (I do not believe even this much, but I will not argue against the thought here.) It is unlikely, however, that we are required to comply with any demands of closure—that we would be making a mistake if we did not.

I suspect that some will disagree with me on this point. Rather than attempting to refute any potential proposal in favor of closure requirements, I will explain how my view accommodates a closure-like phenomenon. Although I am skeptical about closure requirements *per se*, I believe that there are specific cases in which rationality does find fault in those who do not believe the consequences of their beliefs. It is important to be able to account for these cases within the framework I am offering. For example, imagine that you believe that your partner has been receiving calls from an unknown sultry-voiced person, and you also believe that if your partner has been receiving calls from such a person then it is very likely that your partner is having an affair. And suppose that your mind is fixed on these

issues. Many would be inclined to say that you would be irrational if you did not at the same time believe the (perceived) implication: that your partner is very likely having an affair.

Despite appearances, GR4 has something to say about such cases. But to see how the story works, I must first attend to a different part of structural rationality. I will return briefly to closure requirements at the end of the next section.

5.6 – Enkratic Requirements

It is widely accepted that there are requirements of rationality governing enkrasia. In Chapter 3, we saw that even the error theories put forward by the reasons fundamentalists seem to turn on the requirements of enkrasia. Although I do not think that enkrasia is especially central to an account of rationality, I do believe that there are enkratic requirements. Since I am attempting to show that every requirement of structural rationality is rooted in the General Requirement, my goal in this section is to say something about how we might derive enkratic requirements.

There are many related requirements governing the relationships between our beliefs about our normative circumstances and our other attitudes. The phrase “our beliefs about our normative circumstances” is meant to capture judgments like the belief that we have decisive reason to have an attitude, the belief that we lack sufficient reason to have an attitude, the belief that the evidence favors a certain proposition, and so on. To keep things manageable, I will focus on one rather narrow principle of enkrasia, which I introduced in Chapter 1. At the same time, I am confident that what I will say about this enkratic requirement could be extended to deal with other forms of enkrasia. Recall the following principle governing enkrasia in intention:

EI

You are rationally required not to [[believe that you have decisive reason to intend to X] and [not intend to X]].⁷²

According to this requirement, it would be irrational to at the same time believe that you have decisive reason to have an intention and yet not have the intention. The requirement demands that we not suffer from a particular form of akrasia in intention.

What does GR4 have to say about someone who flouts EI? At first glance, it might seem that the answer is “nothing.” Suppose that you believe that you have decisive reason to intend to exercise this evening, and yet you do not have an intention to exercise this evening. GR4 appears to be silent about your rationality. After all, the truth of your belief is not tethered to the intention in question. Your belief is not guaranteed to be false due to the presence or absence of any other attitudes.

⁷² There are two important facts about EI that are worth flagging, and which I have so far ignored. First, my description of the relevant belief is admittedly somewhat artificial. We rarely think in terms of having (or not having) “decisive reason” to intend to do something. EI applies to any agent who has a belief with content equivalent to “I have decisive reason to intend to X.” Second, EI applies only to agents who believe that their so-called *object-given* reasons demand that they X. An example will help to explain why this must be so. Suppose I believe that I have decisive reason to intend to square the circle, but that I have this normative judgment only because I have been told that my intending to square the circle would win me a million dollars. Intuitively, I am not irrational if I fail to have the unrealizable intention. Although you might feel sorry for me for being unable to earn the million dollars, it is unlikely that you would deem me irrational. The important feature of the case is that I would not see myself as having any object-given reasons to intend to square the circle (i.e., my judgment would not be based any considerations having to do with *squaring the circle*). Instead, I would see myself as having tremendous *state-given* reasons to have the intention (i.e., due to the great benefits of being in the relevant mental state). EI does not apply these kinds of cases. This constraint agrees with what I say below concerning the function of self-regarding normative beliefs. Self-regarding normative beliefs based on state-given considerations appear not to have the same function as normal self-regarding normative beliefs.

Nevertheless, I think that we may be able to use GR4 to derive this requirement. What is called for here is not a revision of GR4, but a change in our view about the relevant standard—or standards—of success for the beliefs in question. In the preceding paragraph, I assumed that your belief functionally fails only if it is false. This is a natural assumption since, as I proposed in Section 3, truth is the standard of functional success for belief. But I submit that truth is the sole standard of functional success only across a certain range of beliefs that aim to represent the world outside of ourselves. *Self-regarding normative beliefs*—such as beliefs concerning what we have decisive reason to intend—have an additional standard of functional success, because they play a crucially distinct role in the agency of rational beings. Standard beliefs about the world aspire to represent the world as it is and they succeed just in case they do so correctly. But self-regarding normative beliefs of the sort involved in EI are different. Self-regarding normative beliefs not only aspire to represent the world, but also to bring one’s attitudes into conformity with a perceived normative demand. Self-regarding normative beliefs successfully play their distinctive functional role when they are accompanied by whatever attitude they represent as to-be-held.⁷³ When a self-regarding normative belief is functionally successful, there is no gap at all between the formation of the normative verdict and the holding of the attitude demanded by the verdict.⁷⁴

⁷³ Likewise, a normative belief that holds a particular attitude as not-to-be-held (i.e., the belief that one has insufficient reason to intend to X) plays its distinctive functional role when it directly terminates the relevant attitude (if it was previously held) or ensures the non-generation of that attitude.

⁷⁴ I am not offering a full-blown theory of the relationship between self-regarding normative beliefs and other attitudes. I am only committed to two crucial claims. First, it is important to my view that self-regarding normative judgments (e.g., the belief *that I have decisive reason to intend to flip the*

Of course, the relationship between the normative verdict and the relevant intention can break down. This is what happens in many cases of *akrasia*. You may judge that you have decisive reason to intend to exercise right now but for some reason not also have the intention to exercise right now. There are many possible explanations of how this might happen. I will not attempt to offer an account of the mechanisms of *akrasia* in intention. My point is not to explain how an akratic agent becomes or remains akratic. My point is just that an akratic agent of this sort—one who believes that he has decisive reason to have a certain intention and yet does not have the intention—has a functionally failing self-regarding normative attitude.

It is worth making a few clarificatory remarks. First, I am not suggesting that self-regarding normative beliefs do not also have truth as a functional standard of success. My proposal is that these beliefs play double-duty, in both aiming to represent the world as it (normatively) is and aiming to bring one's attitudes into conformity with one's judgment. My account does not require that attitudes have only a single dimension of functional success and failure (though most do). Second, although I am using the phrase "self-regarding normative beliefs," I am targeting only a select range of such beliefs. For example, a belief *that in June of 2018 I will have decisive reason to try to run a marathon* is a sort of self-regarding normative belief, but because of its time-indexed nature it does not function like the self-regarding normative beliefs I have in mind. A future-oriented self-regarding normative belief like this is similar to a normative belief that I might have about another person. For example,

switch) and the attitudes they concern (e.g., the intention to flip the switch) are distinct attitudes which can come apart, and which do come apart in defective cases. Second, I am committed to self-regarding normative beliefs being functionally involved in bringing about the attitudes that they represent as to-be-held. Each of these claims could be fleshed out in a variety of different ways.

I might believe that Greg has decisive reason to intend to buy a parachute. These normative beliefs do not have any special role to play in our agency. But these beliefs can be converted into beliefs of the kind I have in mind. If I come to believe that now is June of 2018, or come to believe that I am Greg, then I will have a self-regarding normative belief whose functional success depends on my also having the attitudes isolated in the belief.

I am offering only a suggestive sketch of a proposal. Admittedly, a great deal more would need to be said to defend this proposal about self-regarding normative beliefs. But suppose that my proposal is correct: beliefs succeed when they are true, intentions succeed when they realize their objects, and self-regarding normative beliefs (of the sort I have in mind) succeed when they are accompanied by whatever attitude they represent as to-be-held. If this is all true, then GR4 rather easily delivers EI. Suppose that you believe that you have decisive reason to intend to exercise right now, and yet you do not have the intention to exercise. In this case, your normative belief fails (and is guaranteed to fail) as long as you do not have the intention. Enkrasia in intention is prohibited by the General Requirement.

I can now return to the issue of doxastic closure. In Section 5.5, I proposed that there are no structural requirements of doxastic closure *per se*. But I also noted that there seem to be some cases in which failing to believe something obviously implied by other things that you believe is irrational. Recall the example involving your partner's mysterious phone calls. It seems irrational, I proposed, to believe that your partner has been receiving the phone calls, and believe that if your partner has been receiving such phone calls then they are likely having an affair, without also believing that your partner is likely having an affair. I propose that such cases are governed by a requirement of belief enkrasia, rather than a requirement of strict closure. This requirement could be stated as follows:

EB

You are rationally required not to [[believe that you have decisive reason to

believe P] and [not believe P]].

EB can be derived from GR4 in the same way that EI is derived.

We are inclined to view you as irrational in the phone call example because we take it for granted that your network of attitudes includes the judgment that you have decisive reason to believe that your partner is likely having an affair. If you do have a belief of this sort, then we need only to appeal to a requirement of structural rationality governing enkrasia in belief, rather than to a principle of doxastic closure *per se*. So while we cannot derive any strict closure requirements from the General Requirement, the General Requirement allows us to account for the cases which might have otherwise made us inclined to endorse a principle of closure.⁷⁵

5.7 – Structural Rationality and Self-Undermining

I admit that the arguments of this chapter are not exhaustive. There are some putative requirements of rationality which I have ignored entirely. And my proposed derivations and arguments in this chapter are surely open to criticisms that I have not considered, and rely on premises that I have insufficiently defended. But it is beyond my ambitions here to decisively settle all objections. It is enough for my purposes if I have offered a promising and plausible new way of thinking about the foundations of rationality.

I have sketched a theory structural rationality which sees the entire domain of

⁷⁵ This is one point of overlap between my view and the error theories explored in Chapter 3. Although I do not think that we can use the requirements of enkrasia to derive or explain away other requirements of rationality, I think the demands of enkrasia undergird our intuitions about some closure-involving cases.

requirements as rooted in a fundamental prohibition on a certain form of attitudinal self-undermining. I believe that the whole of structural rationality is contained in the General Requirement. To be structurally rational, we must not attitudinally undermine the functional success of our own attitudes.

We are still left with a question: why are the requirements of rationality normative? According to the core argument of this project, answering this question requires that I explain why it is that structural irrationality is a criticizable agential failure. In light of the new finding that irrationality is a matter of self-undermining, it should appear less mysterious why, and in what sense, we ought to be rational. In the next chapter, I will say more about the form of criticizable failure rooted in agents' attitudinal self-undermining.

6.1 – Attitudinal Integrity and Self-Undermining

The normativity of structural rationality has seemed puzzling because insufficient attention has been paid to what rationality, as a whole, is about. By attending to the question of structural rationality’s content—by revealing that rationality’s demands are different applications of the General Requirement of Structural Rationality—it is easier to see why we are bound to comply with its demands. In this brief closing chapter, I will aim to detail the form of criticizable failure involved in violating the General Requirement and in doing so to give an account of the distinctive normativity of structural rationality.

The standard set by the General Requirement is something we might call the *standard of attitudinal integrity*. An agent exhibits attitudinal integrity, in this formal sense, when her intentions and beliefs form a network of attitudes in which each attitude is afforded (by the network) the possibility of success.⁷⁶ In other words, an agent exhibits attitudinal integrity when her network of attitudes is not self-impairing. Structural rationality demands that we meet this standard of attitudinal integrity. Put in negative terms, we can say that structural rationality demands that we not harbor self-undermining collections of attitudes. I propose that a deviation from this standard is a criticizable agential failure and, as such, that the ‘ought’-claims or requirements issued by this standard are normative.

In Chapter 4, I argued that whether deviation from a standard is a criticizable failure

⁷⁶ This means not that each attitude in the network actually *can* succeed (since it may be, for example, that one holds a belief in a necessarily false proposition), but rather that each attitude can succeed *as far as the network of attitudes is concerned*.

(a failure of the normatively relevant sort) is a substantive issue that can be settled only by reflecting on why we might care about our deviation from that standard. So we might ask: “Why care about attitudinal integrity?” or “Why care whether we have self-undermining combinations of intentions and beliefs?” Now, it seems almost undeniable to me that, as a being whose engagement with the world hinges crucially on your beliefs and intentions, you are making a mistake if your employment of your attitudes cuts itself off at the knees. But there is no method by which I can prove my conclusion here. The best I can do is to explain why it seems to me that attitudinal integrity is something that we should (and perhaps could not help but) be concerned with.

Consider the fact that the central, defining feature of our agency is our capacity for belief and intention. Were it not for our ability to represent the world to ourselves and to try to change the world, we would not be agents at all. No doubt, our agency extends both prior to and beyond our intentions and beliefs. Our intentions and beliefs are often (though not always) the culmination of reasoning, and that reasoning is part of our agency. And, on the other side, our intentions and beliefs often (though not always) reach out into the world through our actions, and movements, and words. These external behaviors are parts of our agency, too. I am not denying that there is more to our agency than intending and believing. But intending and believing are, I think, the central and defining aspects of our agency. A being that engaged in reasoning or deliberation but who could not herself hold beliefs or intentions would be unrecognizable as an agent. While reasoning is an important part of our agency, it is not, in itself, what makes us agents. At the same time, an agent whose physiological impairment tragically prevented her beliefs and intentions from extending into the world and resulting in actions would be no less of an agent than the rest of us (though what she could *do* with her agency would be, sadly, limited). Intending and believing is at the

center of our status as agents. We are, above all else, intenders-and-believers.

A structurally irrational agent, then, is an agent who is turned against herself, inhibiting her functioning as the type of being who she is. The problem with the irrational agent is not just that some of her attitudes will fail or, even, that some of her attitudes are certain to fail. Having a failing attitude (e.g., an eternally unrealized intention) or an attitude which is bound to fail (e.g., a belief in a contradiction) is not necessarily a failure attributable to the agent, because things outside of a person can conspire against her to make her attitudes fail. A storm might prevent one of your intentions from being realized, or someone's lie might lead you to have a false belief, or the metaphysical or nomological structure of the world might (unbeknownst to you) leave you with a belief in a complex contradiction. In cases like these, an agent's attitudinal failure does not belong to the agent in the sense I have in mind. For the structurally irrational agent, on the other hand, she alone (rather than anything outside of her) is responsible for ensuring that not all of her attitudes can succeed. My proposal is that this breakdown in her attitudinal integrity undergirds our judgments about rationality and irrationality.

Following Raz, some writers have wondered whether a functional-defect account of irrationality could capture the normativity or bindingness of rational requirements, as I am suggesting. Even if I am correct that every case of irrationality is a case of self-undermining, is this enough to explain the normative force that rational requirements seem to exhibit? After all, the human organism has other dimensions of proper functioning and these are not all tied to normative requirements.⁷⁷ A human agent may be said to malfunction in a certain respect if she has a leaky heart or if she has a physiological condition that inhibits the

⁷⁷ As I suggested in Chapter 3, I believe that this thought is beneath Raz's contention that there are no normative requirements of rationality at all. See Bratman 1987 and Scanlon 2007.

functioning of her appetite or her desire to reproduce. And it would be ridiculous to suggest that there are normative requirements demanding non-leaky hearts or normally functioning appetitive systems. This observation might make us doubt that identifying the form of faulty functioning involved in irrationality is enough to provide an explanation of rationality's normativity.

Of course, I am not suggesting that any old kind of functional deficiency found anywhere inside or relating to an agent can give rise to normative requirements. Attitudinal self-undermining is not much like having a malfunctioning organ or an unusual appetite or reproductive drive. Based on what I have argued above, we can see at least two crucial differences. First, our organs and appetites are dissimilar to intentions and beliefs because intentions and beliefs (but not organs and appetites) are central features of our agency and identity. A malfunctioning in one of my organs or in one of my biological faculties is not a functional failure in *me*, as an agent or person, in any deep sense. Second, an agent's malfunctioning organ or an agent's abnormal biological drive is not attributable to the agent in the way that attitudinal self-undermining seems attributable to the agent. An agent can do things which might contribute to her heart's leaky valve, but only indirectly and contingently. In the case of attitudinal self-undermining, however, the agent is the unmediated cause of her own failure. These two considerations suggest that a breakdown in attitudinal integrity is a special form of functional defect, both because it involves the agent actively and directly *impairing herself* and because it involves an impairment of the agent *as an agent*. This is why irrationality is a criticizable form of failure and the basis for normative claims, while other functional deficiencies may not be.

Here is an easy, though indirect, way of clarifying the point I am pursuing. Upon reflecting on the requirements of structural rationality defended in Chapter 5, some people

might become skeptical that any such requirements exist. Their skepticism would not be rooted in the difficulty of explaining where these requirements come from, but instead in the thought that these principles seem impossible to violate in the first place. These skeptics find the notion of structural irrationality implausible, because they find unintelligible the thought that there could be genuine beliefs and intentions arranged in the relevant ways. When told of the requirement I have called “IC,” for example, a skeptic might say: “But if I were to intend to X, and I really believed that I could not X if I also Y-ed, then I just *could not* really intend to Y while still intending to X. I don’t mean that I *shouldn’t* at the same time intend to Y. I mean that intending to Y is something that I really could not do. To *actually* start intending to Y, I would have to not intend to X. Or I would have to stop believing in the incompatibility. Or something. Having all of those attitudes at the same time makes no sense—I just couldn’t do it.” The skeptic is left thinking that the search for the requirements of rationality is pointless, because they are not principles we could violate.

This skeptical position is too strong. We are, as a matter of fact, sometimes irrational. But the thought fueling the skeptic’s objection is an important one and it is closely related to why structural irrationality is a criticizable failure. The skeptic observes that an agent’s structural irrationality is difficult to interpret, as it is not the behavior of an agent who is behaving as an agent. This is what tempts the skeptic into concluding that some of the agent’s attitudes must not be the genuine article. However, this is the wrong conclusion to draw. The skeptic is correct that what the irrational person is doing “makes no sense” and may border on unintelligibility. This is because the irrational person is an agential mess who is putting her attitudes to work in a way that prevents their success. Structural irrationality is not impossible, but for an intending and believing being there are few clearer grounds for

criticism.⁷⁸ Someone in the business of intending and believing could not help but find fault in their own self-undermining. If someone alerted to their structural irrationality responded with a shrug of the shoulders, we could only conclude that they had not understood what they had been told.

I am running the risk of making irrationality seem much too serious and grave. There is usually no great drama to be found in cases of inconsistent beliefs or intentions, or in other forms of irrationality. I am not suggesting that our attitudinal integrity is some precious condition to be respected, or promoted, or preserved at all costs. (In fact, I am not making any claim at all about the practical importance or instrumental value of avoiding self-undermining.) What I am trying to do is to uncover the basis of the intuitive judgments that led me into this project in the first place. And underneath those judgments, I am suggesting, is the conviction that an agent who guarantees his own attitudinal failure is making a significant mistake, given the type of being who he is.

In virtue of the considerations canvassed here, and in combination with the arguments of Chapter 4, I am persuaded that the requirements of structural rationality are normative in their own right, even though we have no general reasons for complying with them. They are normative because not living up to rationality's standard of attitudinal integrity, as expressed in the General Requirement, is a criticizable failure. We are free, then, to say that you *ought* to comply with rationality's demands, as long as we are aware of what

⁷⁸ In line with the arguments of Chapter 4, it is worth observing that the considerations outlined here will constitute reasons for *caring* about whether or not we comply with the requirements of structural rationality. But this is not an indication that the normativity of the requirements is rooted in reasons. The requirements of structural rationality need not be constituted by, nor offer any, reasons for compliance in order for them to be binding. We are bound to comply with their demands because if we do not then we will be engaged in a form of criticizable failure.

we mean by ‘ought.’⁷⁹

6.2 – The Distinctive ‘Ought’ of Structural Rationality

This conclusion comports with the common-sense judgments I observed at the outset of this project. Suppose I find you in a state of typical structural irrationality, holding a pair of intentions which you believe cannot both be realized. Recall, for instance, case (1) from the opening chapter, in which you intend to spend the weekend with your family, intend to complete your new manuscript on this same weekend, and also believe that doing both of these things together is impossible. There are (at least) two very different kinds of claims that I would make about you and your circumstances. First, I would have some judgment about whether you have reason to maintain or abandon some of your attitudes. I might judge that you ought to focus on your manuscript or instead that you ought to see your family. Or I might judge that you ought to abandon both projects and head out of town to a conference. Or I might point out that you ought to abandon your belief in the incompatibility of your

⁷⁹ It is worth considering my arguments in the light of Southwood 2008. Although Southwood does not explicitly endorse normative pluralism, I believe that his ambition is similar to mine. Southwood argues that it is a mistake to look for a “justification for rational compliance outside of rationality, a source of normativity that is external to rationality” or, in other words, to look for independent reasons to obey rational requirements. There is, at least, a terminological discrepancy between us, as Southwood suggests that in vindicating the normativity of rationality we will locate “reasons that are internal to rationality.” It is difficult to tell whether Southwood is positing that we have so-called intrinsic reasons to be rational, or whether he is offering a use of the word “reason” that is disconnected from the reasons fundamentalists’ usage. I prefer to avoid this use of “reasons” altogether and to say instead that the normativity of rationality is not constituted by reasons. It is worth noting that Southwood’s positive proposal about the unique normativity of rationality is quite different from mine, although there are areas of overlap.

intentions. In any of these judgments, I would be using the ‘ought’ of reasons, which you are bound by in virtue of your status as a reasons-responsive agent. Second, I would judge that, whatever else you do, you ought not have your current combination of intentions and belief. I would see it as a mistake to deploy your attitudes in this self-undermining way, and so I would judge that you ought to do otherwise. Here I would be invoking the ‘ought’ of rationality. I would not be commenting at all on what your reasons demand of you. I would be saying only that in virtue of being an intending and believing agent (and apart from being a reasons-responsive one) it is failure to put your attitudes to work in a way that undermines your agency. This same normative bifurcation is also the best explanation of our judgments about the Mary-Mary* case in Chapter 3, which the reasons fundamentalists were unable to resolve.

If what I am saying is true, then the ‘ought’ of rationality is normatively distinct from the ‘ought’ of most or decisive reason, because they are made relative to fundamentally different standards. This means, as I argued in Chapter 4, that these two different ‘ought’-s do not issue competing kinds of demands. What your reasons tell you to do is one thing and what rationality tells you to do is another. There is not, on the view I am proposing, any shared or ultimate normative measure between reasons and rationality.

I want to bring out both the intuitively satisfying and potentially startling implications of this position. Recall Section 1.3 of Chapter 1, where I considered what we should say about someone who could bring about some tremendous amount of good simply by violating a requirement of rationality. That discussion was meant to show that rationality’s demands cannot simply be the demands of our reasons. Intuitively, an agent who brings about some tremendous amount of good by violating a requirement like IC is, while doing something quite right in one sense, nevertheless making a rational mistake. I pointed out in

Chapter 1 that if we ought to have our attitudes arranged in compliance with IC simply because we have strong reasons to have our attitudes arranged in the relevant way, then we would be unable to explain our judgments about cases in which our reasons favor *not* complying with IC. We now have an explanation of what is going on in these interesting cases. Plausibly, an agent who could bring about some tremendous good by violating IC ought to do so in one sense (i.e., the reasons-involving sense), even though she also ought not to do so in another sense (i.e., the rationality-involving sense). This is an excellent result, because we are of two minds about these sorts of case. We are of two minds about these cases precisely because there are two different normative standards at play.⁸⁰

This result is intuitively appealing. But it may seem to have jarring consequences if my point is misunderstood. An implication of what I am arguing is that there is no ultimate, all-in normative fact about what the agent in this example *really* ought to do, which takes into account both the claims of her reasons and the claims of rationality. This may be an unappealing conclusion, because we may feel inclined to say that (rationality be damned!) the

⁸⁰ Imagine a case in which an agent can rescue a dying person but only by shaving a couple of days off of his own life. Suppose the agent chooses (rightly!) to save the dying person's life. In a case like this, involving competing moral and prudential considerations, we are left with the judgment that the agent has made no normatively relevant mistake *at all* in choosing to save the life. We are not of two minds about cases like this—we will not say “Well, he got something right in one sense, but something wrong in another sense.” Instead, we will judge that the agent has not failed in any normatively salient way. This is because these moral and prudential claims are made relative to the same ultimate standard (i.e., the reasons standard). The case of an agent who violates IC in order to bring about some tremendous value seems very different. Although the agent's violation of IC is praiseworthy in one way, it is obviously a mistake in another way. The fact that the agent brings about a lot of good by violating IC does not cut against or diminish the judgment that violating IC is a mistake. This is strong evidence that the ‘ought’ of rationality is made relative to a non-reasons-based standard.

agent really ought to do the thing that will bring about the tremendous amount of good. If there is an opportunity to do a lot of good, then who cares about rationality? It sounds plain silly to suggest that rationality and reasons are on a par with each other—that either choice would be satisfactory. The demands of our reasons seem much more important than the demands of rationality. My claim that the ‘ought’ of reasons and the ‘ought’ of rationality are fundamentally normatively distinct seems blind to this fact.

I feel the pull of the concern, but it rests on a mistake. I must emphasize that I am *not* saying that rational requirements and the demands of our reasons are on a par with each other. Rationality and reasons could be on a par with each other only if rationality and reasons had some common measure or some shared standard. And this is precisely what I am denying. The demands of rationality and the demands of our reasons are simply different kinds of demands, made relative to fundamentally different and unrelated standards. When we judge that the agent who has the opportunity to bring about a great amount of good simply by violating a requirement of rationality ought to violate the requirement of rationality, we are making this judgment relative to the standard of most or decisive reason. And of course, it goes without saying, relative to *that* standard it will be true that the agent ought to do whatever she can to violate the requirement of rationality. On my view, this is a perfectly legitimate judgment. My point is that the reasons standard is not the only legitimate normative standard. If we were to ask rationality what it thought, it would have a different verdict. Rationality, which sets its sights only on formal (rather than substantive) matters, would demand that you not harbor a self-undermining combination of attitudes. My claim about the separateness of rationality and reasons seems counterintuitive only if it is misunderstood as the claim that the requirements of rationality and reasons are on a par with each other, or for the claim that we would not be making any mistake in complying with IC

and foregoing the tremendous amount of good we might have brought about.

As I indicated in Chapter 4 but could not prove until now, the normative domain as a whole is fragmented into at least two parts. It may seem odd that I am content to reach this conclusion, given that I have been focused on finding unity within the (sub-)domain of rationality. If unity within rationality is so important, then why is fragmentation within the normative domain as a whole acceptable? There is not much to say, I think, other than that this is just the way things are. Rationality has the appearance of a contained class of requirements, while the normative domain as a whole seems broken in half. I did not set out to find fragmentation within the normative domain; there is no sense in defending a version of normative pluralism just for its own sake. But what I have found in this dissertation is that there is compelling evidence that the normative domain is fragmented, that this appearance of fragmentation cannot be explained away or accounted for within a monist system, and that we have access to both a metanormative theory and a theory of rationality which together make sense of all of our intuitive judgments.⁸¹

⁸¹ In light of this conclusion, I now have the opportunity to address a concern that has been lurking in the background since Chapter 1. The concern is that I may have cast the net too narrowly when looking for reasons to comply with the requirements of rationality in the first place. In light of the arguments of Section 6.1, one might ask whether all I have done is uncover the basis of our reasons to comply with rationality's demands.

In Chapter 1, I proposed that our reasons to comply with the dictates of rationality would (if they existed) come in the form of values or benefits that our compliance would promote or serve. I considered, as examples, the possibilities that our compliance with rational requirements might help us to gain more true beliefs, or gain fewer false beliefs, or achieve more of our ends, or produce more total good in the world. But a reader might have worried about my focus on reasons with this instrumental foundation. Not all reasons have this structure. For example, consider my reasons for loving my wife. The fact that it would make me feel happy and fulfilled is (or gives me) some reason to love my wife. This is a reason of the normal, instrumental variety. But, arguably, I also have

6.3 – The Truth in Raz and Korsgaard

As my argument comes to an end, I should pause for a moment to reflect on the extent to which my account of structural rationality mirrors or depends on the arguments of some of the authors I have focused on in this project. In making the arguments of Section 6.1, it would be hard not to notice the extent to which my view overlaps with both Raz's and

reasons to love my wife which have nothing to do with bringing about anything good, simply because loving my wife is something intrinsically worth doing. If there are so-called "intrinsic" reasons to love someone, there may be intrinsic reasons for other things, including intrinsic reasons for complying with rationality's demands. See Kolodny 2005.

We might wonder, then, whether what I have done over the past two chapters is uncover our intrinsic reasons for being rational. According to a possible reinterpretation of my arguments, our reason(s) for complying with the requirements of rationality are grounded in the fact that non-compliance involves an agential breakdown that is intrinsically worth avoiding. It may be that non-compliance has no general, negative consequences (and that compliance has no general, positive benefits). But we could nevertheless have intrinsic reason to comply with rationality's demands, simply because attitudinal integrity is worth preserving in its own right.

It strikes me as awfully fantastical that we have reasons to be rational grounded in the fact that it is intrinsically good to exhibit attitudinal integrity. (If this were the case, then it seems that we would have at least some reason to exhibit this virtue whenever possible; perhaps we should go around collecting as many new sets of consistent beliefs and intentions as we can!) However, if this proposal were true, it would not be such a bad outcome for my argument. It would be a decent achievement to have found these intrinsic reasons that have so far eluded us. And if there were any intrinsic reasons to be rational, the proposal just sketched is as plausible as any.

In light of my arguments of this section, however, I am deeply skeptical that we should rush toward the conclusion that rationality's normativity is constituted by these intrinsic reasons. My conclusion that the normative domain is fragmented is not just a consequence or theoretical commitment of my view, but one of the key benefits of the position. If we were to argue that the normativity of rationality were constituted by some intrinsic reasons to be rational, we would not only be gaining nothing (since we already have a perfectly good, independent explanation of rationality's normativity), but also giving up something important.

Korsgaard's. Indeed, in personal conversations about the view I am putting forward here, I have been asked both whether I am just offering some modification to Korsgaard's constitutivism and whether I am just rehashing Raz's arguments from nearly a decade ago. This is a bit remarkable, given the tremendous space that separates Korsgaard and Raz on the topics of rationality and normativity. So it would be remiss not to say something about the relationship between my view and each of theirs.

To begin, recall this crucial comment from Raz:

The fallacy to avoid is the thought that irrationality [...] occurs only if one fails to conform to a reason [...]. It can consist in faulty functioning, that is in ways of [...] forming beliefs or intentions, and so on, which do not conform to standards of rationality.⁸²

This is very similar to something that I would be happy to write. I agree with Raz that irrationality is not a matter of failing to conform to a reason. (That has been one of the central claims of my project.) And I would also say that irrationality is a matter of faulty functioning in our beliefs and intentions. The difference comes in how narrowly Raz seems to view the notion of faulty functioning, and in what conclusion Raz draws from the fact that we do not have reason to comply with rationality's demands. For Raz, the relevant form of faulty functioning is enkratic failure and the implication of rationality not being based in reasons is that rationality is not a source of normative requirements.⁸³ I depart from Raz's

⁸² Raz 2005a, p. 15.

⁸³ There are moments in Raz's work where it looks as though he may be open to the possibility that rationality involves more than responsiveness to one's perceived reasons. For example, he says that irrationality may be a matter of failing "to conform to rational standards of deliberation, of belief or intention formation, of coherence of belief and intention or others." In isolation, this comment makes Raz seem open to the possibility that rationality is about much more than reasons-responsiveness. But, as far as I am aware, Raz does not follow up on this possibility.

position on both of these points.

Now consider Korsgaard's claim that the constitutive standards of agency deliver unconditionally binding requirements. As I said above about Raz's crucial remark, Korsgaard's claim is something that I might be happy to say myself. Over the last two chapters, I have proposed that as (unavoidable) intenders-and-believers, we are required not to use our intentions and beliefs to thwart the function of those same attitudes. What this means is that the nature of our agency delivers standards which it is a failure not to live up to, even if there is no external purpose served by our meeting those standards. And this is precisely what Korsgaard has argued, even though the content of her arguments is considerably different. I disagree with Korsgaard about a number of crucial points—most importantly about how much we can get out of the conceptual analysis of action, about the form that rational requirements take, and about the role that rationality plays in our lives—but my arguments are indebted to her insights about the possible foundations of normative claims.

It would be misleading to say that my view is a simple combination of Raz's and Korsgaard's or of reasons fundamentalism and constitutivism. But what I have done, I believe, is use a Korsgaardian insight—the idea that we can be bound to live up to the standards set by the nature of our agency—in order to take Raz's original ideas in a different, more positive, and more explanatorily complete direction. As we saw in Chapter 3, the Razian theorists are left appealing to a range of rational requirements (the requirements of *enkrasia*) whose normativity they feel compelled to deny. These error theorists feel that they must deny the normativity of these requirements, because they do not believe that the notion of proper functioning could serve as the basis for normative claims. My arguments over the last three chapters are intended to allay this concern. By adopting the thesis of normative

pluralism, and by unpacking the form of functional failure involved in every instance of structural irrationality, we can accept that irrationality need not involve a failure with respect to our reasons without abandoning the idea that rationality is normative.

6.4 – Conclusions

I should emphasize that the arguments of this dissertation are intended to be conciliatory and hospitable to reasons fundamentalists. Although I have claimed that the reasons fundamentalists are wrong to say that every normative requirement is based in reasons, I think that they would lose very little in giving up on this idea. My proposal does not need to unsettle anything important in their view. We can still hold that there are irreducibly normative facts about reasons, that assessing and responding to reasons is an important aspect of our practical and theoretical agency, and even that the dominant part of the normative domain as a whole is the part having to do with reasons. None of this is lost. But by accepting that the normative domain is broader than the domain of reasons, we gain the ability to do justice to our common-sense judgments about rationality. By accepting my amendment, the reasons fundamentalist gains a lot and loses very little.

In closing, it is worth reflecting on the fact that this dissertation has traveled a great distance to establish a point that may now seem obvious. What I have argued is that structural irrationality is a matter of undermining the success of our own attitudes, that this kind of self-undermining is a mistake, and that this is enough to establish that we ought not to violate the requirements of rationality. Looking back at the starting-point examples from the opening of Chapter 1, I am convinced that this is right. *Of course* when you intend to do two things this weekend while believing you cannot do both you are making it so that you will fail somewhere in your attitudes. And *of course* this is a mistake. And *of course* that makes it

sensible to say that you ought not to hold all of those attitudes at the same time, in some sense of 'ought.' Viewed in a certain light, these conclusions now look almost trivial.

If my conclusions do seem trivial or obvious, I am satisfied with the result. I am content if my project amounts to a long-winded vindication of common sense thinking about rationality. My concern with rationality in the first place was that philosophy was having trouble making sense of a range of my settled intuitions. If it now turns out that those intuitions can be accommodated without damaging anything else of importance, then I have reached my goal.

WORKS CITED

- Bratman, Michael. "Intention and Means-End Reasoning." *Philosophical Review* 90 (1981).
- . *Intention, Plans, and Practical Reason*. Harvard University Press, 1987.
- Broome, John. "Normative Requirements." *Ratio* 12 (1999).
- . "Normative Practical Reasoning." *Proceedings of the Aristotelian Society*. Suppl. 75 (2001).
- . "Practical Reasoning." In José Bermúdez and Alan Millar (eds.), *Reason and Nature: Essays in the Theory of Rationality*. Oxford University Press, 2002.
- . "Have We Reason to Do as Rationality Requires?" *Journal of Ethics & Social Philosophy* 1 (2005).
- . "Wide or Narrow Scope?" *Mind* 116 (2007).
- . *Rationality Through Reasoning*. Wiley-Blackwell, 2013.
- Brunero, John. "The Scope of Rational Requirements." *Philosophical Quarterly* 60 (2010).
- Harman, Gilbert. "Practical Reasoning." *Review of Metaphysics* 29 (1976).
- Kant, Immanuel. Complete works compiled in Paul Guyer and Allan Wood (Eds.) *The Cambridge Edition of the Works of Immanuel Kant*. Cambridge University Press, 1992.
- Kelly, Thomas. "The Rationality of Belief and Some Other Propositional Attitudes." *Philosophical Studies* 110 (2002).
- Kolodny, Niko. "Why Be Rational?" *Mind* 114 (2005).
- . "State or Process Requirements?" *Mind* 116 (2007).
- . "The Myth of Practical Consistency." *European Journal of Philosophy* 16 (2008a).
- . "Why Be Disposed to Be Coherent?" *Ethics* 118 (2008b).
- Korsgaard, Christine. *The Constitution of Agency*. Oxford University Press, 2009a.
- . *Self-Constitution*. Oxford University Press, 2009b.
- . "The Activity of Reason." *Proceedings and Addresses of the American Philosophical Association* 83 (2009c).
- McNaughton, David. "An Unconnected Heap of Duties?" *Philosophical Quarterly* 46 (1996).

- Parfit, Derek. "Rationality and Reasons." In Dan Egonsson et al (eds.), *Exploring Practical Philosophy*. Ashgate, 2001.
- . *On What Matters: Volume 1*. Oxford University Press, 2011a.
- . *On What Matters: Volume 2*. Oxford University Press, 2011b.
- Rabinowicz, Wlodek and Toni Rønnow-Rasmussen. "The Strike of the Demon: On Fitting Pro-Attitudes and Value." *Ethics* 114 (2004).
- Raz, Joseph. *Practical Reason and Norms*. Princeton, New Jersey: Princeton University Press, 1990.
- . *Engaging Reason*. New York: Oxford University Press, 1999.
- . *Value, Respect, and Attachment*. Cambridge: Cambridge University Press, 2001.
- . "The Myth of Instrumental Rationality." *Journal of Ethics & Social Philosophy* 1 (2005a).
- . "Instrumental Rationality: A Reprise." *Journal of Ethics & Social Philosophy* 1 (2005b).
- . *From Normativity to Responsibility*. Oxford University Press, 2012.
- Ross, W.D. *The Right and the Good*. Oxford: Oxford University Press, 2007. Originally published in 1930.
- Reisner, Andrew. "The Possibility of Pragmatic Reasons for Belief and the Wrong Kind of Reasons Problem." *Philosophical Studies* 145 (2009).
- Scanlon, T.M. *What We Owe To Each Other*. Harvard University Press, 1998.
- . "Structural Irrationality." In G. Brennan, R. Goodin, and M. Smith (eds.), *Common Minds: Themes from the Philosophy of Philip Pettit*. Oxford: Oxford University Press, 2007.
- . *Being Realistic about Reasons*. Oxford University Press, 2014.
- Schroeder, Mark. "Means-End Coherence, Stringency, and Subjective Reasons." *Philosophical Studies* 143 (2009).
- . "The Ubiquity of State-Given Reasons." *Ethics* 122 (2012).
- . "State-Given Reasons: Prevalent, If Not Ubiquitous." *Ethics* 124 (2013).
- . "What Makes Reasons Sufficient?" Forthcoming in *American Philosophical Quarterly*.
- Setiya, Kieran. "Cognitivism about Instrumental Reason." *Ethics* 117 (2007).

Sobel, David. "Instrumental Rationality: Not Yet Dead." *Journal of Ethics & Social Philosophy* 1 (2005).

Southwood, Nicholas. "Vindicating the Normativity of Rationality." *Ethics* 119 (2008).

Wallace, R. Jay. "Normativity, Commitment, and Instrumental Reason." *Philosophers' Imprint* 1 (2001).

Way, Jonathan. "Reasons and Rationality." Forthcoming in Daniel Star (ed.) *Oxford Handbook of Reasons and Normativity*.

Williams, Bernard. *Problems of the Self*. Cambridge University Press, 1973.